

(43) International Publication Date 29 November 2001 (29.11.2001)

PCT

(10) International Publication Number WO 01/90354 A1

(51) International Patent Classification⁷: C12N 15/12, 15/85, 15/86, C07K 14/47, 16/18, C12Q 1/68, G01N 33/577, A61K 31/713, 38/18, A01K 67/027

(21) International Application Number: PCT/GB01/02240

(22) International Filing Date: 21 May 2001 (21.05.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

0012186.3

20 May 2000 (20.05.2000) GB

(71) Applicant (for all designated States except US): UNIVER-SITY OF LEEDS [GB/GB]; Leeds LS2 9JT (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): MARKHAM, Alexander, Fred [GB/GB]; Molecular Medicine Unit, Clinical Sciences Building, St. James's University Hospital, Leeds LS9 7TF (GB). JACKSON, Andrew, Peter [GB/GB]; Molecular Medicine Unit, Clinical Sciences Building, St. James's University Hospital, Leeds LS9 7TF (GB). WOODS, Christopher, Geoffrey [GB/GB]; Molecular Medicine Unit, Clinical Sciences Building, St. James's University Hospital, Leeds LS9 7TF (GB).

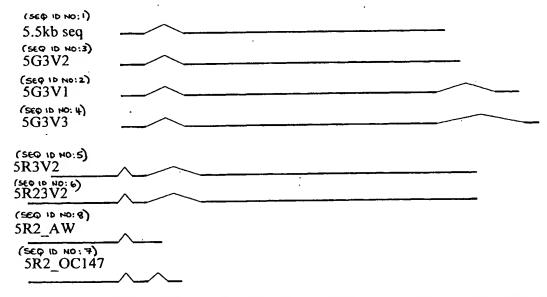
- (74) Agent: HARRISON GODDARD FOOTE; Belgrave Hall, Belgrave Street, Leeds LS2 8DD (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TREATMENT OF CANCER AND NEUROLOGICAL DISEASES



(57) Abstract: The present invention relates to a nucleic acid molecule and the protein encoded thereby absence of which is associated with oral and other cancers and lack of neurogenesis. The invention also provides antibodies and the use of these products as therapeutic and/or diagnostic agents in gene therapy and/or tissue repair.

1/90354 A1

Treatment of Cancer and Neurological Diseases

The present invention relates to the isolation of a nucleic acid molecule and the protein encoded thereby; antibodies raised thereto and the use of these products as therapeutic and/or diagnostic agents particularly, but not exclusively, in gene therapy and/or tissue repair such as, without limitation enhancing neuronal repair /regeneration and in the treatment of cancer.

Background to the Invention

10

15

20

25

30

100

`5'

Oral cancer has significant morbidity and mortality rates. In England and Wales the 5-year survival is around 50%. Globally, oral cancer is one of most common cancers and in some parts of the world it is the most prevalent of all cancer types. For example, in India and Sri Lanka oral cancer accounts for up to 40% of all diagnosed cancers. In addition to geographic "hot spots", there seems to be a rising trend in the increased incidence of oral cancers in many developed nations.

Recent advances in cancer management have failed to impact significantly on the outcome of oral cancer. Surgery and radiotherapy remain the principle forms of treatment with a limited role for chemotherapy. Treatment can be mutilating and is associated with high morbidity that significantly impacts on the quality of life. Speech, swallowing and taste can be markedly impaired after treatment. New treatment modalities are required for oral cancer therapy.

Statement of the Invention

We have identified a gene, from human chromosome 8p23, which is deleted in oral cancer. The gene was found to have distant similarity to the gene encoding the protein "tolloid"; and contains multiple Sushi and CUB domains. We believe that this gene may have utility in diagnosis and gene therapy applications for oral and other cancers.

::::::

5

10

Moreover, and surprisingly, the gene from human chromosome 8p23 may also be implicated in aspects of the developmental regulation of neurogenesis. We base this belief on our observations that the gene has similarity with tolloid, an important developmental gene, and the fact that it is located in the autosomal recessive microcephaly locus, MCPH1, critical region. Sequence variations in this gene can segregate with microcephaly in some families. It therefore may have utility in the diagnosis and therapy of microcephaly, as well as therapies directed to neuronal repair and regeneration, including those utilising stem cells/neural progenitor cells. Having identified this gene we believe that a further use is in the production of transgenic animals. These may have an increased predisposition to oral cancer and/or have decreased or potentially increased neocortex. Such animals would be useful not only as models of oral cancer for the evaluation of novel therapeutics but also to improve understanding of neurological developmental abnormalities. They would also serve as models to test novel therapeutics for neuronal regeneration.

15

20

25

30

According to a first aspect of the present invention there is provided an isolated nucleic acid selected from the group consisting of:

- (a) DNA having the nucleotide sequence given herein as any one of SEQ ID NOS:1 TO 8;
- (b) nucleic acids which hybridize to DNA of (a) above (e.g., under stringent conditions);
- (c) nucleic acids having between 75-95% homology with any one of the nucleotide sequences given herein as SEQ ID NOS:1 to 8; and
- (d) nucleic acids which differ from the DNA of (a), (b) or (c) above due to the degeneracy of the genetic code.

DNAs of the present invention include those coding for proteins homologous to, and having essentially the same biological properties as, the proteins disclosed herein, and particularly the DNA disclosed herein as any one of SEQ ID NOS:1 to 8 and encoding the proteins given herein as SEQ ID NOS:9 to 16 This definition is intended to encompass natural allelic variations therein. Thus, isolated DNA or

cloned genes of the present invention can be of any species of origin, including mouse, rat, rabbit, cat, porcine, and human, but are preferably of-mammalian origin. Thus, DNAs which hybridize to DNA disclosed herein as any one of SEQ ID NOS:1 to 8 (or fragments or derivatives thereof which serve as hybridization probes as discussed below) and which code on expression for a protein of the present invention (e.g., a protein according to any one of SEQ ID NOS: 9 to 16), i.e. the protein lack of which is associated with oral or other cancers and/or lack of neurogenesis of the present invention are to be included in the definition.

5

10

15

20

25

30

Conditions which will permit other DNAs which code on expression for a protein of the present invention to hybridize to the DNAs of SEQ ID NO:1 to 8 disclosed herein can be determined in accordance with known techniques. For example, hybridization of such sequences may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions (e.g., conditions represented by a wash stringency of 35-40% Formamide with 5x Denhardt's solution, 0.5% SDS and 1x SSPE at 37°C; conditions represented by a wash stringency of 40-45% Formamide with 5x Denhardt's solution, 0.5% SDS, and 1x SSPE at 42°C; and conditions represented by a wash stringency of 50% Formamide with 5x Denhardt's solution, 0.5% SDS and 1x SSPE at 42°C, respectively) to DNAs of SEQ ID NO:1 to 8 disclosed herein in a standard hybridization assay. See, e.g., J. Sambrook et al., Molecular Cloning, A Laboratory Manual (2d Ed. 1989) (Cold Spring Harbor Laboratory). In general, sequences which code for proteins of the present invention and which hybridize to the DNAs of SEQ ID NO:1 to 8 disclosed herein will be at least preferably 75% homologous, 85% homologous, and even 95% homologous or more with SEQ ID NO:1 to 8. Further, DNAs which code for proteins of the present invention, or DNAs which hybridize to that given as any one of SEQ ID NOS:1 to 8, but which differ in codon sequence from SEQ ID NO:1 to 8 due to the degeneracy of the genetic code, are also an aspect of this invention. The degeneracy of the genetic code, which allows different nucleic acid sequences to code for the same protein or peptide, is well known in the literature. See, e.g., U.S. Patent No. 4,757,006 to Toole et al. at Col. 2, Table 1.

According to a yet further aspect of the invention there is provided a nucleic acid molecule which encodes a protein lack of which is associated with oral or other cancers and/or lack of neurogenesis and comprises a nucleotide sequence which hybridises to the nucleic acid of any one of SEQ ID NOS:1 to 8 under high stringency conditions.

Preferably, hybridisation occurs under stringent conditions such as 1 x SSC, 0.1% SDS at 65 °C.

Preferably, the nucleic acid is mammalian in origin, for example it may be human or murine.

5

15

20

25

30

Preferably, the nucleic acid of the present invention is at least 2kb and up to 12 kb and may be, for example 5.5kb. The nucleic acid being located on chromosome 8p23.

According to a yet further aspect of the invention there is provided use of the nucleic acid of the present invention, in determining loss of genomic material or loss of expression of mRNA in selected target tissue(s) for diagnosing oral or other cancers and/or neurological developmental abnormalities.

According to a yet further aspect of the invention there is provided use of the nucleic acids of the present invention, in determining the presence of mutants in the DNA and thus diagnosing patients suffering from oral or other cancers and/or neurological developmental abnormalities.

According to a further aspect of the invention there is provided a polypeptide, or a protein comprising an epitope for an antibody or a protein modified by one or more amino acid modifications and comprising an epitope, or a fragment modified or unmodified comprising an epitope for a protein lack of which is associated with oral

or other cancers and/or neurogenesis and encoded by SEQ ID NO:9 to 16. Ideally the polypeptide is encoded by the nucleic acid molecule of any one of SEQ ID NO:1 to 8.

: :::::

10

15

20

25

According to a yet further aspect of the invention there is provided a polypeptide or protein encoded by the nucleic acids of the present invention, preferably the sequences of which are as set forth in SEQID NOS:9 to 16.

According to a yet further aspect of the invention there is provided a delivery vehicle comprising the isolated nucleic acid molecule or polypeptide or protein of the present invention or antibodies to these.

Reference herein to the term delivery vehicle is intended to include any vector whether a viral vector or otherwise for example, without limitation, an adenovirus, a retrovirus, a herpesvirus, a plasmid, a phage, a phagemid or a liposome.

Ideally said delivery vehicle is adapted for administration, for example, but without limitation, by suitable formulation into a suspension.

More preferably, said delivery vehicle is adapted to deliver said nucleic acid molecule or polypeptide to selected tissue. Thus the delivery vehicle is provided with means to facilitate its binding and/or penetration to a specific target site. The nature of the means comprises conventional technologies well known to those skilled in the art for example, without limitation, in the instance where the delivery vehicle is a viral vector said viral vector is provided with surface protein adapted to ensure the viral vector binds to and/or penetrates specific target tissues. Alternatively, gene expression of any one of SEQ ID NOS:1 to 8 may be under the control of a tissue specific promoter. Thus, in this way, the nucleic acid molecule or peptide, fragments or derivatives thereof of the invention can be used in gene therapy treatments.

: :::::

10

15

20

25

According to a yet further aspect of the invention there is provided antibodies raised against the polypeptide, fragment or derivative thereof, of the invention. Ideally the antibodies are monoclonal and more ideally genetically engineered to be humanised. It will be apparent to those skilled in the art that the antibodies of the invention can be used to determine the expression of the polypeptide of the invention in selected target tissue and thus aid in the diagnosis of patients suffering from oral cancers and/or neurological disorders.

According to a yet further aspect of the invention there is provided use of antibodies, fragments or derivatives thereof in diagnosis/detection/identification of oral or other cancers and/or neurological disorders. It will be appreciated that the antibodies as well as the fragments or derivatives of the antibodies recognise the epitope and are capable of binding to the antigenic protein. Also useful are recombinant antibodies. The invention also includes antibodies and other compositions of matter which are specific binding partners of the polyamino acids of the present invention. Reference herein to polyamino acids is intended to include proteins and polypeptides.

The invention further provides for assays using the antibodies of the present invention to detect individuals suffering from or having a predisposition towards oral or other cancers and/or neurological disorders. The assays may employ labelling, for example radioactive labels, enzymes, fluorescent compounds, chemiluminescent compounds, bioluminescent compounds and metal chelates.

Typical assays include assays known to the skilled person for quantitative or non-quantitative detection of antibodies and all involve contacting antigenic polypeptides of the present invention with a sample. The assay may involve for example and without limitation any one or more of the following techniques, RIA, EIA, ELISA, sandwich assays.

According to a yet further aspect of the invention there is provided a method for the treatment of oral cancers and/or neurological disorders comprising administering to a

patient suffering from these conditions the nucleic acid molecule or polypeptide/protein of the present invention.

Preferably, the nucleic acid molecule and/or polypeptide/protein is administered by the incorporation of said nucleic acid molecule or polypeptide/protein into a delivery vehicle as herein described and ideally the method of treatment involves the use of gene therapy.

5

10

15

20

25

30

According to a yet further aspect of the invention there is the nucleic acid and/or protein, as herein before described for use as a pharmaceutical.

According to a yet further aspect of the invention there is provided use of the nucleic acid and/or protein of the present invention for the manufacture of a medicament for the treatment of oral or other cancers and/or neurological disorders.

According to a yet further aspect of the invention there is provided a method of producing a transgenic non-human animal comprising disrupting a gene, or the effective part thereof, the gene comprising the nucleic acid of the present invention and/or the protein or effective part thereof of the present invention.

Reference herein to disruption is intended to include complete or partial disruption of expression of the protein such that the transgenic animal is unable to express levels of the said protein that are typically found in normal individuals as compared with those suffering from oral cancer and/or neurological developmental abnormalities.

Preferably, the transgenic mammal is a rodent and ideally a mouse and more preferably the gene encoding the protein lack of which is associated with oral cancer and/or neurogenesis is the nucleic acid molecule or fragment or derivative thereof as set forth in any one of SEQ ID NOS:1 to 8.

According to a yet further aspect of the invention there is provided a transgenic nonhuman animal whose somatic and germ cells do not contain or express a gene
encoding a nucleic acid, or a nucleic acid which hybridises under high stringency
conditions to, the sequence as set forth in any one of SEQ ID NOS:1' to 8, the gene
having been deleted, mutated or disrupted in the animal or an ancestor of the animal
at an embryonic stage and wherein the gene may be operably linked to an inducible
promoter element.

Preferably, the transgenic mammal is a rodent and ideally a mouse.

10

According to a yet further aspect of the invention there is provided a reporter gene construct based on the promoter region of the gene, or effective part thereof, encoded by any one of SEQ ID NOS:1 to 8 i.e. the nucleic acid of the present invention.

- According to a yet further aspect of the invention there is provided use of a reporter gene construct based on the promoter region of a gene, or effective part thereof, encoded by any one of SEQ ID NOS:1 to 8 in the detection/screening of pharmaceuticals and/or other compounds.
- According to a yet further aspect of the invention there is provided a method of determining the presence of or predisposition towards oral or other cancers and/or neurological developmental abnormalities comprising:
 - (i) identifying the regions of said DNA sample that contain the nucleic acid according to the present invention;
 - (ii) individually hybridising parallel samples of said DNAs with oligonucleotides specific for alleles of the gene encoding any one of said nucleic acids; and
 - (iii) identifying from among said DNA samples those with a loss of heterozygosity for said alleles, wherein identification of a DNA sample with a loss of heterozygosity indicates presence or a predisposition towards neurological developmental abnormalities.

30

Preferably, the DNA sample is obtained from a human patient, alternatively RNA samples may be obtained and used in the method.

Preferably, step (i) may involve amplification of the DNA regions, typically amplification is by PCR.

Brief Description of the Figures

The invention will now be described by way of example only with reference to the following Figures wherein:

Figure 1 represents haplotypes for nine markers from 8p22-pter, for families 1 and 2 segregating autosomal recessive microcephaly. Unaffected siblings from family 1 have been omitted, for clarity. Marker order and relative distances are presented here as deduced from the Généthon map: D8S504-3cM-D8S1824-3cM-D8S1798-3cM-D8S277-2cM-D8S1819-5cM-D8S1825-13cM-D8S552-5cM-D8S1731-5cM-D8S261.

Figure 2 represents sequenced BAC's in this region from the human genome project.

Position of candidate gene sequences 5R-3V2 (SEQ ID NO:5) and 5G-3V2 (SEQ ID NO:3) shown in blue (numbering corresponding to base-pair position in sequence). Sequenced BACs shown in red. BAC clone contig of [Sun, 1999 #387] shown in black, and STSs derived from this contig shown mapped onto the sequenced BACs by the vertical dashed black lines

25

10

15

Figure 3 represents the relationship between SEQ ID NO:1 and the sequence variants of SEQ ID NOS:2 to 8 (not to scale).

SEQ ID NO:1 to 8 represent the nucleic acids of the present invention.

30

SEQ ID NOS: 9 to 16 represent the corresponding protein sequences.

5

10

15

Materials and Methods

Subjects and Methods

A family containing five individuals affected with primary autosomal recessive microcephaly was ascertained. The family originated from the Mirpur region of Pakistan (Fig. 1, family 1). According to the clinical histories, the family confirmed that microcephaly was present from birth in all affected individuals and that there was no history of epilepsy in affected individuals. On examination, head circumferences were 5-9 SD below the population age-related mean. The affected individuals examined were 13-28 years old, and mental retardation ranged from mild to moderate in severity. None were able to read or write, but all could speak and had basic self-care skills. Except for microcephaly, there were no dysmorphic features. No affected individual had a sloping forehead, such as that described by Penrose (Cowie 1960), examination did not reveal weakness, spasticity or athertosis. Computed tomography had been performed on one affected individual at 5 years of age and results were normal. No environmental causes of microcephaly were identified. All parents appeared to be of normal intelligence and had normal head circumferences.

20

A further eight multiply affected consanguineous families were ascertained, with a total of 23 affected individuals displaying primary microcephaly. All of these families also originated from the Mirpur region of Pakistan and had pedigrees consistent with autosomal recessive inheritance.

25

30

DNA Extraction and Microsatellite Analysis

DNA was extracted from peripheral blood lymphocytes by means of a standard nonorganic extraction procedure. The ABI Prism linkage mapping primer set was used to perform a genomewide search. This panel contains 358 microsatellite repeat markers spaced at ~10-cM intervals, with an average heterozygosity of 0.81. PCR amplification of all the autosomal markers was performed according to the

manufacturer's specifications. Amplified markers were pooled and electrophoresed on the ABI Prism 377 gene sequencer with a 4.2% polyacrylamide gel at 3000 V and 52°C for 2 h. Fragment-length analysis was performed using the ABI Prism Genescan and Genotyper 1.1.1 analysis packages.

. .5

100

٠. .

For fine mapping on 8p22-pter, D8S504 and D8S277 from the ABI Prism linkage set were used, and a further seven polymorphic markers from the Genome Database, were selected: tel-D8S1824-D8S1798-D8S1819-D8S1825-D8S552-D8S1731-D8S261-cen. PCR reactions were performed in 10-µl volumes that contained 50 ng genomic DNA; 1µM primers; 250µM each dGTP, dCTP, dTTP, and dATP; 5 U Taq DNA polymerase; and 1 x reaction buffer (1.5-2.0 mM MgCl₂, 10mM Tris-HCl pH 9.0, 50mM KCl, and 0.1% Triton X-100). Amplification was performed with a 5-min initial denaturing step at 95°C; 35 cycles of 94°C for 30 s, 54°C-60°C for 30 s, and 72°C for 30 s; and a final incubation step at 72°C for 5 min.

15

20

25

30

10

Linkage Analysis

A fully penetrant autosomal recessive mode of inheritance was assumed, and the disease allele frequency was estimated at 1/300. Two-point analysis was performed by the LINKAGE analysis programs (Terwilliger and Ott 1994) and HOMOZ-MAPMAKER was used for multipoint anlaysis (Kruglyak et al. 1995). An allele frequency of 0.1 was used in the genome screen for all markers. For further analysis of the candidate region, marker allele frequencies were calculated by genotyping 34 unrelated individuals from the same ethnic population, with a lower limit for allele frequencies set at 0.1. Heterogeneity testing was performed with the HOMOG program (Morton 1955; Terwilliger and Ott 1994).

True Microcephaly was thus mapped to chromosome 8p23 (the MCPH1 locus) (Jackson, 1998) using homozygosity mapping to perform a genomewide search. Refinement of the locus was achieved using further fluorescently labelled primers to microsatellite markers in the region. The overlap between the homozygous regions

from family 1 and 2 (Figure 1) defined the minimal critical region within which the disease gene lies, between D8S1825 and D8S1824. SEQ ID NO 1 maps to this interval on the basis of radiation hybrid mapping data (Genemap 98, Figure 4). This is additionally confirmed from genomic sequence data (SEQ ID NOS: 1 and 9) derived for the gene, which maps the gene to fully sequenced BACs (Figure 2). These BACs map to the critical region by virtue of containing polymorphic markers mapping within the critical region.

Genetic Analysis of Oral Cancers

10

15

20

25

30

: 535

· ·5

Samples of oral cancers were obtained with local Ethics Committee approval from patients undergoing resections of their tumours. DNA was extracted from 20 such tumours and from the corresponding matched normal tissues, by standard techniques well-known in the art, providing 20 pairs of matched normal and oral cancer DNA specimens. Analysis of these paired specimens for loss of particular genetic loci in the tumours, suggestive of the local presence of a tumour suppressor gene, was performed by use of the polymerase chain reaction. Analysis of known microsatellite markers including D8S1806, D8S1824, D8S1781, D8S1788 and D8S262 (see Figure 2) among others, showed frequent loss of one or both alleles at these loci in the majority of the oral tumours. Loss of heterozygosity was particularly frequent at the genetic markers D8S1824, D8S1781 and D8S1788.

The same matched tumour and normal tissue pairs were then compared for alterations in the gene encoding SEQ ID NO:1. In several of these tumours, deletion of both copies of this gene i.e. loss of both alleles, was detected in tumour DNA while PCR products of the expected size were amplified using DNA from matched normal control tissue. In all other cases, the relative amount of PCR amplification product generated using a variety of PCR primer pairs selected within SEQ ID NOS:1 to 8, was markedly reduced in the tumour DNA compared with that generated from normal DNA. In cases where one copy of the gene encoding the SEQ ID NO:1 was apparently retained in tumour tissue, mutations were detected in the remaining DNA

such that the open reading frame encoding the protein of SEQ ID NOS:9 to 16 was disrupted. In every case studied, the change in SEQ ID NOS:1 to 8 resulted in the alteration of a codon encoding a normal amino acid to a mis-sense amino acid or Thus in these cases, the oral cancer cells were unable to termination codon. synthesise the protein of SEQ ID NOS:9 to 16; as a result either of deletion of both copies of the gene described in SEQ ID NOS:1to 8 or as a result of deletion of one copy and truncating or mis-sense mutation in the residual second copy of the gene. This consistent loss of gene expression in tumours is entirely consistent with a role for the protein in SEO ID NOS:9 to 16 as a tumour suppressor protein. It also supports the hypothesis that replacement of a functional gene by provision of the nucleic acid sequence described in SEQ ID NOS:1 to 8 would have therapeutic utility in the treatment of oral and other cancers demonstrating a similar pattern of loss of heterozygosity. Such patterns have been observed in the past for a number of other human malignancies including prostate cancer, breast cancer, ovarian cancer and colorectal cancer. Thus the nucleic acid of SEQ ID NOS:1 to 8 and/or the protein of SEQ ID NOS:9 to 16 may find equal utility in the treatment of these other common human cancers.

Accordingly the nucleic acid molecules and proteins encoded thereby of the present invention and products thereof, are of particular use in gene therapy and in identifying those suffering from or with a predisposition towards cancers, particularly oral cancers and neurological diseases.

5

10

15

References

1. Cowie V (1960). The genetics and sub-classification of microcephaly. J Ment. Defic. Res. 4:42-47.

5

2. Jackson AP, McHale DP, Campbell DA, Jafri H, Rashid Y, Mannan J, Karbani G, Corry P, Levene MI, Mueller RF, Markham AF, Lench NJ, Woods CG (1998). Primary autosomal recessive microcephaly (MCPH1) maps to chromosome 8p22-pter. Am. J. Hum. Genet. 63:541-546.

10

- 3. Morton NE (1955). The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. Am. J. Hum. Genet 7:80-96.
- 4. Terwilliger JD, Ott J (1994). Handbook of human genetic linkage. The Johns Hopkins University Press, Baltimore.
 - 5. Kruglyak L, Daly MJ and Lander ES (1995). Rapid multipart linkage analysis of recessive traits in nuclear families, including homozygosity mapping. Am. J. Hum. Genet. 56:519-527.

20

6. Sun PC, Schmidt AP, Pashima ME, Sunwoo JB and Schlmck SB (1999). Homozygous deletions define a region of 8p23.2 containing a putative tumour suppressor gene. Genomics. 62:184-188.

25

30

P32093wo

Claims

1. An isolated nucleic acid, the nucleic acid being selected from the group consisting of:

: 575:

5

10

- (a) DNAs having the nucleotide sequence given herein as any one of SEQ ID NOS:1 to 8;
 - (b) nucleic acids which hybridise to DNAs of (a) above under stringent conditions;
 - (c) nucleic acids having between 75-95% homology with any one of the nucleotide sequences given herein as SEQ ID NOS:1 to 8; and
 - (d) nucleic acids which differ from the DNA of (a), (b) or (c) above due to the degeneracy of the genetic.
- Nucleic acids according to claim 1 wherein the stringent conditions are 1 x
 SSC, 0.1% SDS at 65 °C.
 - 3. Nucleic acids according to claim 1 consisting essentially of any one of SEQ ID NOS:1 to 8.
- 20 4. Nucleic acids according to claim 1 which hybridise to any one of SEQ ID NOS:1 to 8.
 - 5. Nucleic acids according to claim 1 having between 75-95% homology with any one of the nucleotide sequences given herein as SEQ ID NOS:1 to 8.
 - 6. Nucleic acids according to claim 1 which differ from the DNAs of any one of claims 3 to 5.
- 7. Use of a nucleic acid according to any preceding claim in determining loss of genomic material or loss of expression of mRNA in sample.

8. Use according to claim 7 in detecting the presence of or predisposition towards oral or other cancers and/or neurological developmental abnormalities.

9. Use of a nucleic acid according to any one of claims 1 to 6 in determining the presence of mutants in DNA.

.∙5

₹ ...

15

- 10. Use according to claim 9 in identification of patients suffering from oral or other cancers and/or neurological developmental abnormalities.
- 10 11. A polypeptide or a protein encoded by the nucleic acid molecules of any one of claims 1 to 6.
 - 12. A delivery vehicle comprising any one of the isolated nucleic acid molecules of claims 1 to 6 or the polypeptides or proteins encoded thereby or antibodies to these polypeptides or proteins.
 - 13. A delivery vehicle according to claim 12 comprising a viral vector selected from the group comprising an adenovirus, a retrovirus, a herpesvirus, a plasmid, a phage, a phagemid or a liposome
 - 14. A delivery vehicle according to either claim 12 or 13 provided with surface protein adapted to facilitate binding and/or penetration to a specific target.
- 15. A pharmaceutical composition comprising a nucleic acid according to any one of claims 1 to 6, a polypeptide or protein according to claim 11 and/or the delivery vehicle of any one of claims 12 to 14 and a suitable excipient, diluent or carrier.
- 16. Antibodies which are specific binding partners of the polypeptide/protein of claim 11 or fragment or derivative thereof which are capable of binding to the antigenic part of the polypeptide/protein.

17. Antibodies according to claim 16 which are monoclonal and/or genetically engineered to be humanised.

: :::::

10

- 18. Use of antibodies or antibody fragments according to either claim 16 or 17 in determining the presence or level of expression of the polypeptide or protein of claim

 11.
 - 19. Use of antibodies or antibody fragments according to either claim 16 or 17 or fragments or derivatives thereof in detecting the presence or absence of binding partners whose absence is indicative of oral or other cancers and/or neurological disorders.
 - 20. A method for the treatment of oral cancers and/or neurological disorders comprising administering to a patient suffering from or predisposed to these conditions the nucleic acid molecule of any one of SEQ ID NOS:1 to 8 and/or the proteins encoded thereby.
- 21. A nucleic acid according to any one of claims 1 to 6 or polypeptide or protein of claim 11 or delivery vehicle of any one of claims 12 to 14 for use as a pharmaceutical.
 - 22. A polyamino acid as set forth in any one of SEQ ID NOS: 9-16 for use as a pharmaceutical.
- 25 23. Use of the nucleic acids according to any one of claims 1 to 6, for the manufacture of a medicament for the treatment of oral or other cancers and/or neurological disorders.
- 24. A method of producing a transgenic non-human animal comprising disrupting a gene comprising the nucleic acid of any one of claims 1 to 6, or the effective part

: 5:5:

20

thereof, the gene encoding a protein or effective part thereof lack of which is associated with oral or other cancers and/or lack of neurogenesis.

- 25. A method of producing a transgenic non-human animal comprising preventing expression of a protein or polypeptide of claim 11, or the effective part thereof, lack of expression of the protein being associated with oral or other cancers and/or lack of neurogenesis.
- 26. A transgenic non-human animal whose somatic and germ cells do not contain or express a gene encoding a nucleic acid according to any one of claims 1 to 6, the gene having been deleted, mutated or disrupted in the animal or an ancestor of the animal at an embryonic stage and wherein the gene may be operably linked to an inducible promoter element.
- 15 27. A transgenic non-human animal according to any one of claims 24 to 26 wherein the animal is a rodent.
 - 28. A reporter gene construct based on the promoter region of the gene, or effective part thereof, comprising the nucleic acid of any one of claims 1 to 6.
 - 29. Use of a reporter gene construct based on the promoter region of a gene, or effective part thereof, comprising the nucleic acid of any one of claims 1 to 6 in the detection/screening of pharmaceuticals and/or other compounds.
- 25 30. A method of determining the presence of or predisposition towards oral cancer comprising:
 - (i) identifying regions of a DNA sample that contain the nucleic acid according to any one of claims 1 to 6;
- (ii) individually hybridising parallel samples of said DNAs with
 oligonucleotides specific for alleles of the gene encoding any one of
 said nucleic acids; and

(iii) identifying from among said DNA samples those with a loss of heterozygosity for said alleles, wherein identification of a DNA sample with a loss of heterozygosity indicates presence or a predisposition towards oral cancer.

5

- 31. A modified method according to claim 30 wherein the sample comprises RNA.
- 32. A method of determining the presence of or predisposition towards neurological developmental abnormalities comprising:
 - (i) identifying regions of a DNA sample that contain the nucleic acid according to any one of claims 1 to 6;
 - (ii) individually hybridising parallel samples of said DNAs with oligonucleotides specific for alleles of the gene encoding any one of said nucleic acids; and
 - (iii) identifying from among said DNA samples those with a loss of heterozygosity for said alleles, wherein identification of a DNA sample with a loss of heterozygosity indicates presence or a predisposition towards neurological developmental abnormalities.

20

- 33. A modified method according to claim 32 wherein the sample comprises RNA.
- 34. A kit comprising the nucleic acids of any one of claims 1 to 6 and a set of instructions for use thereof.

5

SEQ ID NO:1

cDNA sequence (partial) 5.5kb

ttttagggatggtatgaatttaatatttttagtattacaatatattcttataaaaaaggtccaaqtq aaaaaggcgattgagttgaagtcaagaggagtcaagatgctgcccagcaaggATGGAAGCCATAAAAA CTCTGTCTGGCATATGGAATAACATCAACCATGTGACATCCGAAGAAGATACGTTCATTATGTATCTG 10 GGAAAACCATGGCTTCAAGTGAAAATTCAAGTGAGCCAAGGAGGTGTTGCATTGGTCTCTGACATGTG TCCAGATCCTGGGATTCCAGAAAATGGTAGAAGAGCAGGTTCCGACTTCAGGGTTGGTGCAAATGTAC AGTTTTCATGTGAGGACAATTACGTGCTCCAGGGATCTAAAAGCATCACCTGTCAGAGAGTTACAGAG 15 TCATCACCACCACCGACCGGACAAGGTÇATCAAGCTTGCCTTTGAAGAGTTTGAGCTGGAGCGAGGC TATGACACCCTGACGGTTGGTGATGCTGGGAAGGTGGGAGACACCAGATCGGTCTTGTACGTGCTCAC GGGATCCAGTGTTCCTGACCTCATTGTGAGCATGAGCAACCAGATGTGGCTACATCTGCAGTCGGATG GGAATCCCCGCCTATGGGAAGCGGACGGGCAGCAGTTTCCTCCATGGAGATACACTCACCTTTGAATG 20 CCCGGCGCCTTTGAGCTGGTGGGGGAGAGAGTTATCACCTGTCAGCAGAACAATCAGTGGTCTGGCA ACAAGCCCAGCTGTGTATTTTCATGTTTCTTCAACTTTACGGCATCATCTGGGATTATTCTGTCACCA **AATTATCCAGAGGAATATGGGAACAACATGAACTGTGTCTGGTTGATTATCTCGGAGCCAGGAAGTCG** AATTCACCTAATCTTTAATGATTTTGATGTTGAGCCTCAATTTGACTTTCTCGCGGTCAAGGATGATG GCATTTCTGACATAACTGTCCTGGGTACTTTTTCTGGCAATGAAGTGCCTTCCCAGCTGGCCAGCAGT 25 GGGCATATAGTTCGCTTGGAATTTCAGTCTGACCATTCCACTACTGGCAGAGGGTTCAACATCACTTA CACCACATTTGGTCAGAATGAGTGCCATGATCCTGGCATTCCTATAAACGGACGACGTTTTGGTGACA GGTTTCTACTCGGGAGCTCGGTTTCTTTCCACTGTGATGATGGCTTTGTCAAGACCCAGGGATCCGAG TCCATTACCTGCATACTGCAAGACGGGAACGTGGTCTGGAGCTCCACCGTGCCCCGCTGTGAAGCTCC ATGTGGTGGACATCTGACAGCGTCCAGCGGAGTCATTTTGCCTCCTGGATGGCCAGGATATTATAAGG 30 ATTCTTTACATTGTGAATGGATAATTGAAGCAAAACCAGGCCACTCTATCAAAATAACTTTTGACAGA CGGCGAGTACCACGGCACCCAGGCACCCCAGTTCCTCATCAGCACCGGGAACTTCATGTACCTGCTAT TCACCACTGACAACAGCCGCTCCAGCATCGGCTTCCTCATCCACTATGAGAGTGTGACGCTTGAGTCG GATTCCTGCCTGGACCCGGGCATCCCTGTGAACGGCCATCGCCACGGTGGAGACTTTGGCATCAGGTC 35 CACAGTGACTTTCAGCTGTGACCCGGGGTACACACTAAGTGACGACGAGCCCCTCGTCTGTGAGAGGA ACCACCAGTGGAACCACGCCTTGCCCAGCTGCGACGCTCTATGTGGAGGCTACATCCAAGGGAAGAGT GGAACAGTCCTTTCTCCTGGGTTTCCAGATTTTTATCCAAACTCTCTAAACTGCACGTGGACCATTGA AGTGTCTCATGGGAAAGGAGTTCAAATGATCTTTCACACCTTTCATCTTGAGAGTTCCCACGACTATT TACTGATCACAGAGGATGGAAGTTTTTCCGAGCCCGTTGCCAGGCTCACCGGGTCGGTGTTGCCTCAT 40 ACGATCAAGGCAGGCCTGTTTGGAAACTTCACTGCCCAGCTTCGGTTTATATCAGACTTCTCAATTTC GTACGAGGGCTTCAATATCACATTTTCAGAATATGACCTGGAGCCATGTGATGATCCTGGAGTCCCTG CCTTCAGCCGAAGAATTGGTTTTCACTTTGGTGTGGGAGACTCTCTGACGTTTTCCTGCTTCCTGGGA GCCAAGGTGTGTGGCCGAATGTGGAGCAAGTGTCAAAGGAAATGAAGGAACATTACTGTCTCCAAATT 45 TTCCATCCAATTATGATAATAACCATGAGTGTATCTATAAAATAGAAACAGAAGCCGGCAAGGGCATC CACCTTAGAACACGAAGCTTCCAGCTGTTTGAAGGAGATACTCTAAAGGTATATGATGGAAAAGACAG TTCCTCACGTCCACTGGGCACGTTCACTAAAAATGAACTTCTGGGGCTGATCCTAAACAGCACATCCA ATCACCTGTGGCTAGAGTTCAACACCAATGGATCTGACACCGACCAAGGTTTTCAACTCACCTATACC AGTTTTGATCTGGTAAAATGTGAGGATCCGGGCATCCCTAACTACGGCTATAGGATCCGTGATGAAGG 50 TGACCTGTTTGAGTGGAGACAGGAGAGTGTGGGACAAACCACTACCTTCGTGCATAGCGGAATGTGGT GGTCAGATCCATGCAGCCACATCAGGACGAATATTGTCCCCTGGCTATCCAGCTCCGTATGACAACAA CCTCCACTGCACCTGGATTATAGAGGCAGACCCAGGAAAGACCATTAGCCTCCATTTCATTGTTTTCG ACACGGAGATGGCTCACGACATCCTCAAGGTCTGGGACGGGCCGGTGGACAGTGACATCCTGCTGAAG 55 CAGCGACTTCTTCATCAGCAAGTCTGGCTTCTCCATCCAGTTCTCCACCTCAATTGCAGCCACCTGTA ACGATCCAGGTATGCCCCAAAATGGCACCCGCTATGGAGACAGCAGAGAGGGCTGGAGACACCGTCACA TTCCAGTGTGACCCTGGCTATCAGCTCCAAGGACAAGCCAAAATCACCTGTGTGCAGCTGAATAACCG GTTCTTTTGGCAACCAGACCCTCCTACATGCATAGCTGCTTGTGGAGGGAATCTGACGGGCCCAGCAG 60 GTGAACCCGGACTTTGTCATCGCCTTGATATTCAAAAGTTTCAACATGGAGCCCAGCTATGACTTCCT

ACACATCTATGAAGGGGAAGATTCCAACAGCCCCCTCATTGGGAGTTACCAGGGCTCTCAGGCCCCAG AAAGAATAGAGAGTAGCGGAAACAGCCTGTTTCTGGCATTTCGGAGTGATGCCTCCGTGGGCCTTTCA GÁCAAGAGTTGGAACAGACTTCAAGCTTGGCTCCACCATCACCTACCAGTGTGACTCTGGCTATAAGA 5 TTCTTGACCCCTCATCCATCACCTGTGTGATTGGGGCTGATGGGAAACCCTCCTGGGACCAAGTGCTG CCCTCCTGCAATGCTCCCTGTGGAGGCCAGTACACGGGATCAGAAGGGGTAGTTTTATCACCAAACTA CCCCCATAATTACACAGCTGGTCAAATATGCCTCTATTCCATCACGGTACCAAAGGAATTCGTGGTCT TTGGACAGTTTGCCTATTTCCAGACAGCCCTGAATGATTTGGCAGAATTATTTGATGGAACCCATGCA CAGGCCAGACTTCTCAGCTCACTCTCGGGGTCTCACTCAGGGGAAACATTGCCCTTGGCTACGTCAAA TCAAATTCTGCTCCGATTCAGTGCAAAGAGCGGTGCCTCTGCCCGCGGCTTCCACTTCGTGTATCAAG 10. CTGTTCCTCGTACCAGTGACACCCAATGCAGCTCTGTCCCCGAGCCCAGATACGGAAGGAGAATTGGT TCTGAGTTTTCTGCCGGCTCCATCGTCCGATTCGAGTGCAACCCGGGATACCTGCTTCAGGGTTCCAC GGCGCTCCACTGCCAGTCCGTGCCCAACGCCTTGGCACAGTGGAACGACACGATCCCCAGCTGTGTGG TACCCTGCAGTGGCAATTTCACTCAACGAAGAGGTACAATCCTGTCCCCCGGCTACCCTGAGCCATAC 15 GGAAACAACTTGAACTGTATATGGAAGATCATAGTTACGGAGGGCTCGGGAATTCAGATCCAAGTGAT CAGTTTTGCCACGGAGCAGAACTGGGACTCCCTTGAGATCCACGATGGTGGGGGATGTGACCGCACCCA CATTTCCAGTCTGACATTAGTGTGGCAGCTGCTGGTTTCCACCTGGAATACAAAACTGTAGGTCTTGC TGCATGCCAAGAACCAGCCCTCCCCAGCAACAGCATCAAAATCGGAGATCGGTACATGGTGAACGACG TGCTCTCCAGTGCGAGCCCGGGTACACCCTGCAGGGCCGTTCCCACATTTCCTGTATGCCAGGG 20 ACCGTTCGCCGTTGGAACTATCCGTCTCCCCTGTGCATTGCAACCTGTGGAGGGACGCTGAGCACCTT GGGTGGTGTGATCCTGAGCCCCGGCTTCCCAGGTTCTTACCCCAACAACTTAGACTGCACCTGGAGGA TCTCATTACCCATCGGCTATGGTGCACATATTCAGTTTCTGAATTTTTCTACCGAAGCTAATCATGAC TTCCTTGAAATTCAAAATGGACCTTACCACACCACCCCATGATTGGACAATTTAGCGGCACGGATCT 25 CCCCGCGGCCTGCTGAGCACAACGCATGAAACCCTCATCCACTTTTATAGTGACCATTCGCAAAACC GGCAAGGATTTAAACTTGCTTACCAAGCCTATGAATTACAGAACTGTCCAGATCCACCCCCATTTCAG AATGGGTACATGATCAACTCGGATTACAGCGTGGGGCAATCAGTATCTTTCGAGTGTTATCCTGGGTA CATTCTAATAGGCCATCCTCCG

SEQ ID NO:2

				210110.2		
	5G-3V1 Nuc	cleotide sec	quence	6145 bp		
: 5:5	; 1	TTTTAGGGAT	GGTATGAATT	TAATATTTTT	TAGTATTACA	ATATATTCTT
	51	ATAAAAAAGG	TCCAAGTGAA	AAAGGCGATT	GAGTTGAAGT	CAAGAGGAGT
5	101	CAAGATGCTG	CCCAGCAAGG	ATGGAAGCCA	TAAAAACTCT	GTCTGGCATA
	151	TGGAATAACA	TCAACCATGT	GACATCCGAA	GAAGATACGT	TCATTATGTA
	201	TCTGGGAAAA	.CCATGGCTTC	AAGTGAAAAT	TCAAGTGAGC	CAAGGAGGTG
	-251	TTGCATTGGT	CTCTGACATG	TGTCCAGATC	CTGGGATTCC	AGAAAATGGT
* ***	4 301	AGAAGAGCAG	GTTCCGACTT	CAGGGTTGGT	GCAAATGTAC	AGTTTTCATG
10	. 351	TGAGGACAAT	TACGTGCTCC	AGGGATCTAA	AAGCATCACC	TGTCAGAGAG
1,5,	401	TTACAGAGAC	GCTCGCTGCT	TGGAGTGACC	ACAGGCCCAT	CTGCCGAGCG
	451	AGAACATGTG	GATCCAATCT	GCGTGGGCCC	AGCGGCGTCA	TTACCTCCCC
	501	TAATTATCCG	GTTCAGTATG	AAGATAATGC	ACACTGTGTG	TGGGTCATCA
	551	CCACCACCGA	CCCGGACAAG	GTCATCAAGC	TTGCCTTNGA	AGAGTTTGAG
15	601	CTGGAGCGAG	GCTATGACAC	CCTNACGGTT	GGTGATGCTG	GGAAGGTGGG
13	· 651	ACACACCACA	TCGGTCTTGT	ANGTGCTCAC	GGGATCCAGT	GTTCCTGACC
	701	TCATTCTCAC	CATGAGCAAC	CACATCTCCC	TACATCTGCA	GTCGGATGAT
	751 751	ACCAPTCCCT	CACCTGGGTT	TABACCTCTT	TACCAAGAAA	TTGDADAGGG
		AGCATIGGCI	GATCCTGGAA	TOCCCCCCTA	TECENACIES	ACCCCCACCA
20	801	AGGGTGTGGG	TGGAGATACA	CHCACCAGCCIA	AATCCCCCCC	CCCCTTTCAC
20	851	GTTTCCTCCA	AGAGAGATACA	CICACCITIG	CACAACAATC.	ACTCCTTTGAG
****	901	CTGGTGGGG	AGAGAGTTAT	CACCIGICAG	CAGAMCAMIC.	AGIGGICIGG
	951		AGCTGTGTAT	TITCATGITI	CITCAACITI	ACGGCATCAI
	1001	CTGGGATTAT	TCTGTCACCA	AATTATCCAG	AGGAATATGG	GAACAACAIG
~-	1051	AACTGTGTCT	GGTTGATTAT	CTCGGAGCCA	GGAAGTCGAA	TTCACCTAAT
25	1101	CTTTAATGAT	TTTGATGTTG	AGCCTCAATT	TGACTTTCTC	GCGGTCAAGG
	1151	ATGATGGCAT	TTCTGACATA	ACTGTCCTGG	GTACTTTTTC	TGGCAATGAA
	1201	GTGCCTTCCC	AGCTGGCCAG	CAGTGGGCAT	ATAGTTCGCT	TGGAATTTCA
	1251	GTCTGACCAT	TCCACTACTG	GCAGAGGGTT	CAACATCACT	TACACCACAT
	1301	TTGGTCAGAA	TGAGTGCCAT	GATCCTGGCA	TTCCTATAAA	CGGACGACGT
30	1351	TTTGGTGACA	GGTTTCTACT	CGGGAGCTCG	GTTTCTTTCC	ACTGTGATGA
	1401	TGGCTTTGTC	AAGACCCAGG	GATCCGAGTC	CATTACCTGC	ATACTGCAAG
	1451	ACGGGAACGT	GGTCTGGAGC	TCCACCGTGC	CCCGCTGTGA	AGCTCCATGT
	1501	GGTGGACATC	TGACAGCGTC	CAGCGGAGTC	ATTTTGCCTC	CTGGATGGCC
	1551	AGGATATTAT	AAGGATTCTT	TACATTGTGA	ATGGATAATT	GAAGCAAAAC
35	1601	CAGGCCACTC	TATCAAAATA	ACTTTTGACA	GATTTCAGAC	AGAGGTCAAT
	1651	TAŢGACACCT	TGGAGGTCAG	AGATGGGCCA	GCCAGTTCGT	CCCCACTGAT
	1701	CGGCGAGTAC	CACGGCACCC	AGGCACCCCA	GTTCCTCATC	AGCACCGGGA
	1751	ACTTCATGTA	CCTGCTATTS.	ACCACTGACA	ACAGCCGCTC	CAGCATCGGC
	1801	TTCCTCATCC	ACTATGAGAG	TGTGACGCTT	GAGTCGGATT	CCTGCCTGGA
40	1851	CCCGGGCATC	CCTGTGAACG	GCCATCGCCA	CGGTGGAGAC	TTTGGCATCA
	1901	GGTCCACAGT	GACTTTCAGC	TGTGACCCGG	GGTACACACT	AAGTGACGAC
	1951	GAGCCCCTCG	TCTGTGAGAG	GAACCACCAG	TGGAACCACG	CCTTGCCCAG
	2001	CTGCGACGCT	CTATGTGGAG	GCTACATCCA	AGGGAAGAGT	GGAACAGTCC
	2051	TTTCTCCTGG	GTTTCCAGAT	TTTTATCCAA	ACTCTCTAAA	CTGCACGTGG
45	2101	ACCATTGAAG	TGTCTCATGG	GAAAGGAGTT	CAAATGATCT	TTCACACCTT
	2151	TCATCTTGAG	AGTTCCCACG	ACTATTTACT	GATCACAGAG	GATGGAAGTT
	2201	TTTCCGAGCC	CGTTGCCAGG	CTCACCGGGT	CGGTGTTGCC	TCATACGATC
	2251	AAGGCAGGCC	TGTTNGGAAA	CTTCACTGCC	CAGCTTCGGT	TTATATCAGA
	2301	CTTCTCAATT	TCGTACGAGG	GCTTCAATAT	CACATTTTCA	GAATATGACC
50	2351	TGGAGCCATG	TGATGATCCT	GGAGTCCCTG	CCTTCAGCCG	AAGAATTGGT
	2401	TTTCACTTTG	GTGTGGGAGA	CTCTCTGACG	TTTTCCTGCT	TCCTGGGATA
	2451	TCCTTTAGAA	GGTGCCACCA	AGCTTACCTG	CCTGGGTGGG	GGCCGCCGTG
	2501	TCTCCACTCC	ACCTCTGCCA	AGGTGTGTGG	CCGAATGTGG	AGCAAGTGTC
	2551	ANACCAAATC	AAGGAACATT	ACTGTCTCCA	AATTTTCCAT	CCAATTATGA
55	2601	ACAGGAAA1G	GAGTGTATCT	ATABABTAGA	AACAGAAGCC	GGCAAGGGCA
22		TANIANCCAI	AACACGAAGC	TTTCCACCTCT	TTCDDCCDCD	TACTCTAAAG
	2651	CONTRACTOR	GAAAAGACAG	TTCCACCIGI	CCACTCCCCA	CCTTCACTA
	2701	GIAIAIGAIG	CTGGGGCTGA	TOCTOROU	CACATCCAAT	CACCTGTGGC
	2751	AAATGAACTT	CACCAATGGA	TCCIMAGCAG	DCCD DCC.d.d.d.	TCAACTCACC
60	2801	TAGAGTTCAA	TTGATCTGGT	TOTARCACAC	CATCCCCCCA	ΨΟΟΟΤΟΛΟΟ
60	2851	TATACCAGTT	ATCCGTGATG	VAUNT OT ONG	TACCCACACA	CAPCALCACTY
	2901	CGGCTATAGG	ATCCGTGATG	CCCVACCAAC	CCACCAACACI	CCTCTCTCTGT
	2951	ACAGTTGCAA	CCCGGGGTAC	CMCCCATGCATG	CCACTACCTT	CCIGACCIGI
	3001	TTGAGTGGAG	ACAGGAGAGT	GTGGGACAAA	CCACTACCTT	CGIGCATAGC

3051							
3151		3051	GGAATGTGGT	GGTCAGATCC	ATGCAGCCAC	ATCAGGACGA	ATATTGTCCC
3151	هر	3101	CTGGCTATCC	AGCTCCGTAT	GACAACAACC	TCCACTGCAC	CTGGATTATA
S		3151					
3251 ACATCCTGCT GAAGGACTGC ACTCACTCCC CCTTCCCGGA GGACATCCCCC 3361 ACACCATCA ACTCACTCAC CCTGACTTC GACAGCACT TCTCACAGC 3461 ACACTCCAGG TATCCCCCAA AATGGACCC GCTATGGAC ACCACAGAGA 3451 GCTGGAGACA CCGTCACATT CCAGTGTGCC GCTATGGAG ACACAGAGAGA 3451 GCTGGAGACA CCGTCACATT CCAGTGTGCC GCTATGGAGA CAGCAGAGAGA 3451 GCAGCCCTCC TACATGCACTA CCAGTGTGCC GCTATGGAGA CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG							
3301 AGCACCTTCA ACTCACTCAC CCTCACTTC CACACCACCTTCA	5						
3351	3						
3401 ACGATCCAGG TATGCCCCAA AATGGCACC GCTATGGAGA CAGCACCAGG 3501 ACAAGCCAAA ATCACCTGTT TCCAGCTGAC CAGCGCTATC TTTTGCAACCAGA 3501 ACAAGCCAAA ATCACCTGTT TCCAGCTGAA TAACCGGTTC TTTTGCAACCAGA 3501 GCAGGTGTAA TTTTGTCACC CAACTACCCA CAGCCGTATC CTCCTGGGAA 3601 GCAGGTGTAA TTTTGTCACC CAACTACCCA CAGCCGTATC CTCCTGGGAA 3601 GCAGGTGTAA TTTTGTCACCA CAACTACCCA CAGCCGTATC CTCCTGGGAA 3601 GAAGGGAAG ATTCCAACAG CCCCCTCATT GGGAGTTACC CACCACTCTAT 3701 TATTCAACAAG CTCCCCTCGATT GGGAGTTACC CAGCCTCTATC 3801 GGCCCAGAA AGAATAAGAG CTCCCCTCATT GGGAGTTACC CAGCCTATC 3851 GGAGGATGAC CTCCCTGGGC CTTTCAGGGAT CAGCCTGTAT CTGGCATTC 3851 GGAGGAAC ACACTCCACC CTCCACCTAC CAGCCTACC 3901 AAACCACGGG AACCTTGTTT TGACCCAGGA ATATAATGA ATTTAAACA 3901 AAACCACGGG AACCTTGTTT TGACCAGGA ATATAATGA ATTGGACCAG 4001 CTGGCTATAA GATCTTGAC CCCTCATCCA TCACCTAC CAGTTGGGCT 4001 CTGGCATATA CACAGCTGGT CAAATATCC CCTCCATCCA ATATAACCA 4001 CTGGAGCCCAG ATCACCGGGA CAGAAGGGT AGGTTTTACC CAGACCTAC 4001 CTGGAGCCCAG ATAATATTC ATTGCAT CACCGTACC ACCAGACCTAC 4001 CTGCACTCTC GGGGTCTCAT CACCGGAAC CACCTCTCA CAGCCTCAC 4001 CTGCACTCTC GGGGTCTCAT CACCGGAAC CACCTCCAC CACCACCTAC 4201 AAGCAATCC GAAATATTTC ATGGAACCA TCCCCCTGC CACCCCGG 4201 AAGCAATCC GAAATATTTC ATGGAACCA TCCCCCTGC CACCCCGG 4301 GCTCACTCTC GGGGTCTCAC TCAGGGAAC CATCCCCTA 4301 GCTCACTCTC GGGGTCTCAC TCAGGGAAC CATCCCCTA 4301 GCTCACTCTC GGGGTCTCAC TCAGGGAAC CATCCCCTGC CACCCCGG 4401 CTTCCACCTC GGGTCTCAC TCCTGCCCC CACCCCTGC TACCAGGCCT GCCCACGCTG 4501 GCCGGCTCCA TCCCCACAC TCCCCACGCCT GCCCCACGCCC CACCCCCGG 4501 CCCAGCCCCAC CCCACGCCCC GACCCCCACGC CCCGGCCCCCACGC CCCACGCCCCACGC CCCACGCCCCGG 4501 CCCAGCCCCCACGCCCCCACGC CCCGCGCCCCCCCCCC							
10 3551 GCTGGAGACA CCGTCACATT CCAGTGTGAC CAGTGCTAC AGCGCCACA 3551 CAGACCCACA ACTACCATA ACTACCACACA 3651 GCAGGCTGCT TATTGTCACC CAGTGCTACACA AGCGGCACACA 3651 GCAGGTGTTA TTTTGTCACC CAGTCACCTACCC CAGCGGCTACC CTCCTGGGAA 3651 GGAATGTGAC TTTTGTCACC CAGCGCTATTC CTCCTGGGAA 3751 TATTCAAAAGA ATTCCAACAGA CCCCCTCATT GGGGGTTACC AGGGCTCTCA 3801 GGCCCCAGAA AGAATAGAGA CTCCATGGGT ATGACTTCCT CACACTTCAT 3851 GGAGTGATGC CTCCTGGGC CTTCTGGGT TCGCCTTCACA 3801 AGACCACGG AACATTGTT TGACCAGGA AATTAAAAGA 3901 AAACCACGGG AACATTGTT TGACCAGGA AATTAAAAGA 3901 AAACCACGGG AACATTGTT TGACCAGGA AATTAAAACA AATTAAACAG AATTAAAACAG AATTAAAACAG AATTAAACAG AATTAACCAGCAGAG CAAAGAGGA AATTAAACAG AATTAA							
3501 ACAAGCCAAA ATCACCTGTT TECAGCTGAA TAACCGGTTC TITTGGGAC		3401	ACGATCCAGG	TATGCCCCAA	AATGGCACCC	GCTATGGAGA	CAGCAGAGAG
3551 CAGACCCTCC TACATGCATA GATGGTTTOTG GAGGGATATCT GACGGGCAA		3451	GCTGGAGACA	CCGTCACATT	CCAGTGTGAC	CCTGGCTATC	AGCTCCAAGG
3551 CAGACCCTCC TACATGCATA GATGGTTTOTG GAGGGATATCT GACGGGCAA	10	3501	ACAAGCCAAA	ATCACCTGTG	TGCAGCTGAA	TAACCGGTTC	TTTTGGCAAC
3601 GCAGGTGTTA TTTTGTCACC CAACTACCCA CAGCGCTTAC CTCCTGCAGCA 3651 GGAARGTGAC TGGAGATCA AAGTCAACC GACTTTGCT ACACATCTAT 3751 GAAGGGAAG ATTCAACAG CCCCCTCATT GGGAGTTACC CACACTCTAT 3851 GGACCCAGAA AGAATAGACA GCACCCTCATT GGGAGTTACC CACACTCTAT 3851 GGACGARGC CTCCGTGGCC CTTTCAGGGT TCGCCATTCA ATTTAAACACA 3951 AAGCCACGGA AACCTCTATT TGACCCAGGA AATTAAACACA 3951 AGATGGAACA GACTTCAACC CTTCGGCATTCA CTTCGGCATTCA CTTCGGCATTCA CTTCGGCATTCA CTTCGGCATTCA CTTCGGCATTCA CACACTCAC CACTCGGCA A051 GATGGGAACA GACTTCAACC CCCTCATCCA CACCTGGGC CACACGCACA CACACCTAC CACACCT	•						
1551 GGARTEGAC TGGAGAGTAN ANGTGANCCC GGACTTTETC ACCCTTGA 1701 TATTCANAMA TTTCANACAG CCCCCAGTA TGGACTCCT ACACATCTAT 1801 GGCCCCAGAN AGANTAGAGA GTACGAGCCAGTA TGGACTCCT ACACATCTAT 1801 GGCACCAGAN AGANTAGAGA GTACGAGGANA CAGCCTGTTA CTGGACTTCA 1801 GGCACCAGAN AGANTAGAGA GTACGAGGANA CAGCCTGTTA CTGGACTTCA 1801 AAACCACGGG AACCTGGTT TGACCCAGGA AATATAATGA ATTGGACACAG 1901 CTGGCTATAA GATTCTAGAC TTGGCTCCAC TCACCCTGT GATTGGGCATAG 1901 CTGGCTATAA GATTCTTGAC CTGGCTCCAC TCACCCTGT GATTGGGCTA 1901 CTGGCTATAA GATTCTTGAC CCCTCATCCA TCACCTGTG GATTGGGCTA 1901 CTGGCTATAA GATTCTTGAC CCCTCATCCA TCACCTGTG GATTGGGCTA 1901 CTGGAGGCAG TACACGGGAT CAGAAGGGT AGTTTTATCA CCAAACTACC 1901 CTGGAGAGCAG TACACGGGAT CAGAAGGGT AGTTTTATCA CCAAACTACC 1901 CTGCACTTC GGGACTCTCCAC TCACACTGC ATGCTCCCCACTGACA 1901 CCCCACTACTC CAGAGGGAT CAGATAGCC TGAAACAGCC AGACTTCCCA 1901 CTCCACTTC GGGGTCTCAC TCAGAGGGAN AGTTTCCCAT CAGAGCACA 1901 GCCCACTCCA CAGACCAGA TACACGGGAN AGTTTCCCAT CAGAGCACA TGCACAGGCC AGACTTCCCA 1901 GCCCACTCCC GGGGCTCACA TCACACGGAA AGCCCATGACA 1901 GCCCGCTCCA TCGTCCATT CAGAGCACA TGCACAGCCC AGACTACCA 1901 GCCCGCTCCA TCGTCCATT CAGAGCACA CCCAATGCA 1901 CTCCACTTCC CTGACCCAGA TACGCAGAGA CACCAATGCA 1901 CTCCACTTCC CCCAGCCCAGA TACGCAGAACAA ACCCAATGCA 1901 CTCCACTTCC CCCAGCCCAGA TACGCAAGACCA ACCCAATGCA 1901 CTCCACTTCC CCCAGCCCAGA TACGCAAGACCA ACCCAATGCA 1901 CTCCACTCTC AGACCACACC CCGGGATAC TGCACCAGGAACAA 1901 CTCCAACTGC AGACCACACA CCCAGGCCCAGA TACGCAAGACAA 1901 CTCCAACTGC AGACCACACA CCCAGGCCCAGA TACGCAACAACAA 1901 CTCCAACTGC AGACCACACA CCCCAGCCCAGA CTCCAACCACACCAAGACCAA 1901 CCCAGACGAT CAGACCACACCACACCACACCACACCAAGACCAA 1901 CCCAGACGAT CAGACCACACCACACCACACCACACCAC							
15 3751 TATTCAAAAG TITCAACATG GAGCCCAGT ATGACTTCT ACCACTCAT 3851 GGCCCCAGAA AGAATAGAGA CTCCCTCATT GGGACTTACC AGGCCTCTAT 3851 GGCCCCAGAA AGAATAGAGA CTAGCCGCAGAA AGATTCAGAGT CTCCTGGGGC CTTTCAGGGT TCGCCATTGA ATTTAAAGAG 3951 AAACCACGGG AAGCTTCTAGC CTTTCAGGGT TCGCCATTGA ATTTAAAGAG 3951 AGATGGAACA GACTTCAAGC TTGGCCACCAC CATCACCTACC CAGTGGGACAAG 4051 CTGGCTATTAA GATTCTTGAC CCCTCATCCA TCACCTGTGC CAGTGGGACAAG 4051 GATGGGAACA CCTCCTGGGA CCAACGGGT CAGCACGAGA AGTTGACCCAC CATCACCTACC CAGTGGGCACAG 4151 CCCATAATTA CACAGCGGAT CAGAGGGGT CAGTTTCCAC ATGCCCCTGC CATCACCCTGCA CAGCCCTGAA CAGCCCATACCCACACACACACACACACACACACACACAC							
155							
3801 GGCCCCAGAA AGAATAGAGA CTAGGGGAAA CAGCCTGTTT CTGGCATTCA	15						
3851 GGAGTGATEC CTCCTGTGGC CTTTCAGGGT TCGCCATTGA ATTTAAAGAG 3951 AGATCACGGG AAGCTTGATT TGACCCAGGA AATATAATGA ATGGGACAG 3951 AGTTGAACA GACTTCAAGC TTGGCTCCAC CATCACCTAC CAGTGTGACT 4051 GATGGGAACA CATCTCTGGC CCCCATCCA TCACCCTGT GATTGGGGCT 4101 TGGAGGCCAG TACACGGGAT CACAGTGCTG CCCTCCTGCA ATGCTCCCTG 4101 TGGAGGCCAG TACACGGGAT CACAGTGCTG CCCTCCTGCA ATGCTCCCTG 4101 TGGAGGCCAG TACACGGGAT CACAGTGCTG CCCTCTCTCA TACCCGTACA 4201 AAGGAATTG TGGTCTTTGG ACAGTTGCC TCTATTCCAT CACGGTACCA 4201 AAGGAATTG TGGTCTTTGG ACAGTTGCC TATTTCCAG CAGCCCTGAA 4301 GCTCACTCTC GGGGTCTCAC TCAGGGGAAA CATTGCCCTT GGCTACCTCA 4351 AATCAAATC TGCTCCAGTT CAGGGGAAA CATTGCCCTT GGCTACCTCA 4351 AATCAAATC TGCTCCGATT CAGGGGAAA CATTGCCCT TGCCCCGGG 4401 CTTCCACTCT CGGGTCTCAC TCAGGGGAAA CACGGTGCC TGCCCCGCGA 4551 TCCACGGGG CTCCACTCCC ATCCTCCC ACGGGAAAC CACCAATGCA 4551 TCCACGGGC CTCCACTCCC ATCCGTCC CAGCCCCAAC CACCAATGCA 4551 TCCACGGGG CTCCACTCCC ATCCGTCC CAGCCCCAAC CACCAATGCA 4661 ACGACACGAT CCCCAGCTCT GGGTACCCT GCAGTGGCAA TTTCACTCAA 4661 ACGACACGAT CACCCAATGCA CCCGGCTAC CAAGCCCTACA ACGGAAACAA 4701 CTTGAACTGT AATATGGAAC TCATAGTTAC GCAGAGTGGAA TTTCACTCAA 4861 CACAGTGGT CAGTTTTGCC ACGGAACCAA ACTGGGAACCA 4861 CACAGTGCT CAGTTTTGCC ACGGAACCAA ACTGGGAACCA 4861 CACAGTGCT CAGTTTTGCC ACGGAACCAA ACTGGGAACCA 4861 CACAGTGCT CATTGTCAC CAGGAACCAA ACTGGGAACCA 4861 CACAGTACCG GGATTCACT CAGGAACCAA ACTGGGAACCA TCTGGGACCT CCTGAGCATC 4861 CACAGTACCG GACTGCTGA ACAGTACTTC CACCCAACCC 4861 CACAGTACCA CATTGTCAC CACGCAACCA ACTGGGAACT TCCCAGCGA 4861 CACAGTACCA CATTGTCAC CACGCAACCA ACTGGGAACCA CACGCAACCA 4861 CACAGTTCCC GGGATCACC CACGCACCACA ACTGGGAACCA CACGCAACCA 4861 CACAGTTCCC GGGATCACC CACGACCACAC ACTGGGAACCA CACCACACCA	13						
3901 AAACCACGGG AAGCTTGTTT TGACCCAGGA AATATAATGA ATGGGACAAGC							
20 4001 CTGGCTATAA GATTCTTGAC CCCTCATCCA CACTGCTAC CAGTGTGACT ACST ACST ACST ACST ACCT ACCT ACCT A		3851	GGAGTGATGC	CTCCGTGGGC	CTTTCAGGGT	TCGCCATTGA	ATTTAAAGAG
4001 CTGGCTATAA GATTCTTGAC CCCTCATCCA TCACCTGTGT GATTGGGCCT 4101 TGGAGGCCAG TACACGGGAT CAGAAGGGGT AGCTTATTCA CCAAACTACC 4151 CCCATAATTA CACACGGGAT CAGAAGGGGT AGTTTTATCA CCAAACTACC 4151 CCCATAATTA CACACGTGGT CAAATATGCC TCATTCCAT CAGGGTACCA 4201 AAGGAATTCG TGGTCTTTGA CACAGTTGCC TATTTCCAGA CAGCCTGAA 4201 AAGGAATTCG TGGTCTTGG ACAGTTGCC TATTTCCAGA CAGCCCTGAA 4301 GCTCACTCT GGGGTCTCAC TCAGGGGAAA CATTGCCCT 4301 GCTCACTCT GGGGTCTCAC TCAGGGGAAA CATTGCCCT 4401 CTTCCACTTC GTGTATCAAG CTGTCACG TACCAGTGCC TCGCCCGGG 4401 CTTCCACTTC GTGTATCAAG CTGTCCAGT TACCAGTGCA ACCCAATGCA 4451 GCTCTGTCCC CAGGCCCAGA TACGGAAGGA GAATTGTT TGGTTTTTT 30 4501 GCCGGCTCA TCGTCCGATT CGAGTGCAAA CACGAATGCA 4551 TTCCACGGC CTCCACTGCC ACTCCCTG TACCAGTGAC ACCCAATGCA 4651 CACAAGACAT CCCCAGCTT GTGGTAACCA CCGGGATACC TCTCTCAGGG 4551 TTCCACGGCG CTCCACTGCC ACTCCTGCC CAACGCCTTG GCACACTGGA 4601 ACGACACCAT CCCCAGCTT GTGGTACCACT GGAGGCCAT ACGGAAACAA 4701 CTTGAACTGT AATTGGAAGA TCATAGTTAC GGAGGGCTC GGAATCAGA 4751 TCCAAGTGAT CAGTTTTGCC ACGGACCAGA ACTGGGAACCA 4851 CACAGTACCG GCACTGCTGA CACGCAGAC ACGGAACCAA 4751 TCCAAGTGAT CAGTTTTGCC ACGGACCAGA ACTGGGACTC TCTCAGGCAC 4851 CACAGTACCG GCACTGCTGA ACGGAACCAA CTGGGAACCA 4751 TCCAAGTGAT CAGTTTTGCC ACGGACCAGA ACTGGGAACCA 4751 TCCAAGTGAT CAGTTTTGCC ACGGACCAGA ACTGGGACTC TACCTGCAT 4901 TCCAGTCTGA CATTAGTGTG GCACCCCAGA CTGGGAACCA TCCCAGCACC 4851 CACAGTACCG GCACTGCTGA ACGGACCAGA CTGGCACTC TACCTGCAT 4901 TCCAGTCTGA CACCTTGCA GCCCCCACCAGA CTGGCACCT TACCTGCAT 4901 TCCAGTCTGA CACCTTGCA GCCCCCGCC CCTCCCCA GCAACACCT 5001 ACCGTCCGC GTTGCAACTA CCCCTGCAG CTGCACCTCCC TTCCACTACCACCT 5011 ACCGTTCGC GTTGCAACTA CCCCTGCAG CTGCACCACC CCCACCACCAC 5011 ACCGTCCCC GTGCAACTA TCCCTCCCC GGCTTCCC ACACACCAC CCCACCACCAC 5011 ACCGTTCCC GCACCACCACCC CCTGCTCCC ACACACCAC 5011 ACCGTTCCC GCACCACCACC CCCTGCTGAG CCCCTCCCA 5011 ACCGTTCCC GCACCACCACC CCCTGCTGAG CACCACACC 5011 ACCGTTCCC GCACCACCACCC CCCTGCTGAG CACCACACC 5011 ACCGTCTCC GCACCACCCC CCCCGCGCGCG CCCCTCCCCA GACCACCCC 5011 ACCGTCCCC TTCCCACC TTCCCCC GGCCACCCCC CCCCCACACCACC 5011 TCAGGCAC TCCCCCGGATT TCCCCCACC	₹.	3901	AAACCACGGG	AAGCTTGTTT	TGACCCAGGA	AATATAATGA	ATGGGACAAG
4001 CTGGCTARTAA GATTCTTGAC CCCTCATCCA TCACCTGTGT GATTGGGGCT 4101 TGGAGGCCAG TACACGGGAT CAGAAGGGGT AGCTTATTCA CCAAACTACC 4151 CCCATAATTA CACACGTGGT CAAATATGCC TCTATTCCAT CAGGGTACC 4151 TGATTGGCA GAATTATTA CACACGTGGT CAAATATGCC TATTTCCATA CAGGGTACC 4201 AAGGAATTGC TGGTCTTTGA CACAGTTGCC TATTTCCAGA CAGCCCTGAA 4201 AAGGAATTGC TGGTCTACAC TCAGGGGAAC CATTGCCCT 4301 GCTCACTCT GGGGTCTCAC TCAGGGGAAC AGCCCTGAA 4351 AATCAAATTC TGCTCCGATT CAGTGCAAAG AGCGGTGCT CTCCCCGCG 4401 CTTCCACTTC GTGTATCAAG CTCTCTCG TACCAGTGCC CTCCCCCGG 4401 CTCCACTTC GTGTATCAAG CTCTCTCG TACCAGTGCA CCCAATGCA 4551 GCTCTGTCCC CGAGCCCAGA TAGGGAAGG GAATTGGTT TGGTTTTTT 30 4501 GCCGGCTCCA TCGTCCGATT CGAGTCCAC CCGGGATACC TCCTCAGGGG 4551 TTCCACGGC CTCCACTGCC AGCCCTGG GAATGCAC CCCAATGCA 4651 CACAAGACGT CCCCAGCTGT GTGGTACCACT GCACTGGGAA 4601 ACGCACCAGT CCCCAGCTGT GTGGTACCAC GGAGGCCTG GAAGGAACAA 4701 CTTGAACTG TATTGGAAGA TCATAGTTAC GGAGGCCTC GGAATCAGA 4751 TCCAAGTGAT CAGTTTTGCC ACGGACCAGA ACTGGGAACCA 4851 CACAGTAGCG GCACTTGCC ACGGACCAGA ACTGGGAACCA 4851 CACAGTACCG GCACTGCTGA CCCCAGGACCAGA ACTGGGAACCA 4851 CACAGTACCG GCACTGCTGA CAGGACCAGA ACTGGGAACCA 4951 ACTGTAGGTC TTGCTCGCAG CCCACCCAGA CTGGGAACCT TCTCAGGCAC 4851 CACAGTACGC GCACTGCTGA CAGGACCAGA ACTGGGAACCA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA CTGCCATCT TACCTGCAT 4901 TCCAGTCGA GATTAGTGT CCAAGAACCA CTGCCATCTC TACCTGCAT 4901 TCCAGTCGA GACCCTTGCA GCCCCCAGC CTGGAACCA TTCCACTCACA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA CTGCCTTCC TACCCTGCAT 5001 CAAAATCGGA GATCGGTACA TGGTGAACCA CTGTCCTCC TTCCACTGCG 5051 AGCCCCGGGTA CACCCTTGCA GCCCCCCCCCCCC GGAATCCAA 4901 CCAGTCTGC GATTGTCA TGTGTGAACA CTGCCTCTCC TTCCACTGCG 5051 AGCCCCGGGTA CACCCTTGCA GCCCCCCCCCCC CCTCCTCCCC GGAATCCAA 5001 CAAAATCGAA GACCCTTTTA TGGTGAACCA CTGCACCACCC CCCACACCAC 5001 CAAAATCGAA ATGGGAACCA TCTCCTCCCC GGCCTCCCCC GGCTTCCCCA 5001 CAAAATCGAA ATGGGAACCA TCTCCTCCCC GGCCTCTCCC AAGCCCATGCC 5001 TCTCACCCC ACCACCACCC CCCCGCGCCCCCCCCCCC	•	3951	AGTTGGAACA	GACTTCAAGC	TTGGCTCCAC	CATCACCTAC	CAGTGTGACT
4051 GATGGGAAC CCTCCTGGGA CCAGATGCTG CCCTCTGCA ATCCTCCTG 4151 CCCATAATTA CACAGCTGGT CAAATATGCC TCTATTCCATC CCAAACTACC 4201 AAGGAATTCG TGGTCTTTGG ACAGTTGCC TCTATTCCATC CAGAGTACCA 4201 AAGGAATTCG TGGTCTTTGG ACAGTTGCC TTATTCCAGC CAGCCCTGAA 4251 TGATTTGGCA GAATTATTTG ATGGAACCA TGCACAGGCC AGACTTCTCA 4301 GCTCACTCT GGGGTCTAC TCAGGGGAAA CATTGCCCTT GGGTACCTA 4351 AATCAAATTC TGCTCCGATT CAGTGCAAA AGCGGTGCCT CTCCCCCGGGG 4401 CTTCCACTCT GTGTATCAAC CTTTCTCATG TACCAGTGCA 4451 GCTCTCCTC GTGTATCAAC CTTTCCTC TACCAGTGCA 4551 TTCCACGGGC CTCCACTCC AGTCCAAGGCA GCCCTGAA 4501 GCCGGCTCCA TCGTCCGATT CCAGTGCAAC GCAGTGCA 4601 ACGACACGAT CCCCAGCTGT GTGGTACCC CAAGGCCTTG GCACAGTGCA 4601 ACGACACGAT CCCCAGCTGT GTGGTACCC CAAGGCCTTG GCACAGTGCA 4601 CCGAAGGGTA CAATCCTGTC CCCCGGCTCC CAAGGCCTTG GCACAGTGCA 4701 CTTGAACTGT ATATGGAACA TCATAGTTAC GGAGGGCTCG GCAAATCCAGA 4701 CTTGAACTGT ATATGGAACA TCATAGTTAC GGAGGGCTCG GGAATTCAGA 4701 CTTGAACTGT ATATGGAACA TCATAGTTAC GGAGGGCTCG GGAATTCAGA 4701 CCTGGATGCG GGAGTGCAC CCCCGGCCCCAG ACTGGGAAGCT TCTCAGGCAC 4801 CACGATGGG GGGATGTGAC CGCACCCAGA ACTGGGAAGCT TCTCAGGCAC 4801 CACGATGGG GGGATGTGAC CGCACCCAGA CTGGGAAGCT TCTCAGGCAC 4801 CACGATGGG GGAGTGTGAC CGCACCCAGA CTGGGAAGCT TCTCAGGCAC 4801 CACGATGGA GACCTTGCA GCAGCCTCC ACATTTCCTC TACCTGCAGC 5001 CAAAATCGGA GATCGGAACA TGGGAACCA TGGGAAACAA 4951 ACCGTTCGC GTTGCATC CAACACACC CCCTCCCCA GCAACACCAT 4901 TCCAGGTCTG ACCCTGCAG GCCCTTCCC CACGATTCCCT TACCTGGACC 5001 CAAAAATCGGA GATCGGTACC TGGTGCTCC TTTCCACTC GCAACACCAT 5001 CACAATTCAC GCCCTCCCC GTGGCTCTC TTTCCACTC GCAACACCACC 5001 CACAATTCAC GCCCTCCCC GGAGGACT CCTCACCCCCAGCCCCC GCACCACCACCCCCCAGCCCCTC CCTGCACCCCC GCACCACCACCCCCCC GCACCACCACCCCCCCC	20	4001					
4101 TGGAGGCCAG TACACGGGAT CACAAGGGGT AGTTTATCA CCAAACTACA 4151 CCCATAATTA CACAGCTGGT CAAATATGC TCTATTCCAT CACAGGTACA 4201 AAGGAATTCG TGGTCTTTG ACAGTTTGCC TATTTCCACA CAGCCTGAA 4201 AAGGAATTCG GGGTCTACT CAGGGGAAA CATTCCCCTT GGCTACCTC GGGTCTCC GGGGTCTCAC TCAGGGGAAA CATTCCCCTT GGCTACCTC GGGTCTCACT CAGGGGAAA CATTCCCCTT GGCTACCTCA 4351 AATCAAATTC TGCTCGATT CAGTGCAAAG CATTCCCCTT GGCTACCTCA 4401 CTTCCACTTC GTGTATCAAG CTGTCCTCA TACCAGTGCA 4401 CTTCCACTTC GTGTATCAAG CTGTCCTCA TACCAGTGCA 4601 GCCGGCCCAGA TACGGAAGAG GAATTGTTC CACCAGTGCA 4451 GCCGGGCTCC TCGCCCCAGA TACCAAGCCTT GCCCAGCTAC 4651 TTCCACGGCG CTCCACTCC CAGCCCAGA TACCAAGCCTT GCACAGGAACA 4651 CCGAAGAGGTA CACTCACCCCAGA CCCAAGCCTTG GCACAGCCTTG GCACAGGCATACA 4651 CCGAAGAGGTA CAATCCTGTC CCCGGCTAC CCTGAGCCAT GCACAGCATACA 4651 CCGAGAGAGTA CAATCCTGTC CCCGGCTAC CCTGAGCCAT GCACAGAACAA 4701 CTTGAACTGA TATTGGAAGA TACTAAGTAC GGAAGCCTC GCACAGCATACAA 4701 CTTGAACTGA TATTGGAAGA TACTAAGTAC GGAAGCCTC GCACACCAGA CTGGAACAA 4851 CACAGTAGCC GCACTCCTCA CACGCACAC CCCTGCAGAACCT TTCCAGGCAC 4851 CACAGTAGCC GCACTCCTCA CACGCACAC CCCTGCCAT 4901 TCCAGTCGC CATTAGTGG GCAGCTGCTG GTTTCCACCT GAAAATCAAA 4951 ACCGTCAGC CACCTCCCAA GACCCATC CCCTGCATC CACAGAACCA TACCAACACAT 4901 TCCAGTCGC GACCCTTGCAG CACCCTAGA CCCTCCCAG GCCCCTCCCA GCACAACCAT 4901 TCCAGTCGC GACCCTTGCAG CACCCTAGA CCCCTCCCAG GCCCCTCCCA GCACACACCA 5001 CACAGTACGAA 6601 CACAGAACCA 6601 CACAGAACACA 6601 CACAGACCA 6601 CACAGAACCA 6601 CACAGACCA 6601 CACAGACCA 6601 CACAGACCA 6601 CACAGACCA 6601 CACAGACCA 6601 CACACCCC 6601 CACACACCA 6601 CACACACCA 6601 CACACACCA 6601 CACACACCA 6601 CACACACCA 6601 CACACACCA 6601	_ •						
4151 CCCATAATTA CACAGCTGGT CAAATATGCC TCTATTCCAG CAGGCTACA 4201 AAGGAATTCG TGGTCTTTGG ACAGTTTGCC TATTTCCAGA CAGCCCTGAA 4301 GCTCACTCTC GGGGTCTCAC TCAGGGGGAAA CAGTTGCCCTT GCCCAGGGC 4301 AATCAAATTC TGCTCCGATT CAGTGCAAAG ACGGTGCCCT GGCCCCGGG 4401 CTTCCACTC GTGTATCAAG CTGTTCCTC TACCAGGCC AGCCCAGGC 4451 GCTCTCTCC CGAGCCCAGA TACGGAAGAA CAGGGTGCCT CTGCCCGCGG 4451 GCCGCTCCA CTGCCCGATT CAGTGCAAAG ACGGGTACC TGCCCGCGG 4551 TTCCACGGC CTCCACTGC TACGGAAAC CGGGATACC TGCTCTCAGGG 4551 TTCCACGGC CTCCACTGC AGTCCGTAC CCGAGTGCAA CGGGATACC TGCTCTCAGGA 4651 CGAACAGGAT CCCCAGCTT GTGGATCCAC CCGGGATACC TGCTCCAGATGCAA 4701 CTGAACTGT ATATGGAAGA TCATAGTTAC GGAGGGCAT ACGGAAACAA 4701 CTGAACTGT ATATGGAAGA TCATAGTTAC GGAGGGCCT CCTTGAGATC 4801 CACGATGGTG GGGATGTAC CCCAGCTTC CCTGAGCCAT ACGGAAACAA 4701 CTCAACTGC GCACTGCCAA ACAGGACCAA ACTGGAACCAA 4801 CACGATGGTG GGCATTGCCC CCCAGCCTA CCTGAGCCAT CCTTGAGATC 4801 CACGATGCTG GCATGCTCA CCGAGCACA ACTGGAACCAA 4801 CACGATGCTG GCATGCTCA CCCAACCACTC TACCTCCATT 4901 TCCACTCTGA CATTAGTGTA CCGACCCAGA CTGGGAAGCCT TCTCACGCAT 4901 TCCACTCTGA CATTAGTGTA CCGAGCACGA CTGGGAAGCCT TCCCACGTT 4901 TCCACTCTGA CATTAGTGTA CCGACCCAGA CTGGGAAGCCT TCCCACGTT 4901 ACCGTTCGC GTTGCACTC CCAAGAACCA GCCCTCCCA GCAATACAAA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAATACAAA 4951 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTTTCCACTGCG 5011 ACCGTTCCC CAACAACTTA GTTTCTGATT TTTTCTACCG AACCTGTGG 5011 TGACTTCCT GAAATTCAA ATGGACCAA CCGAGATCC AACCTGTGG 5011 ACACATTTCC CAACAACTTA TCCACCTC GGAGGATCT CAACACCACA 5011 TGACTTCCT GAAATTCAA ATGGACCAAC CCGACACCAC ACCACACAC 5011 GAAAATTAA CCGACACTTA TAGGAACAA CAGAACCA AACCCCAAATTC 5011 TAAACTTCCT TACCAAGATT TAGGAACAA CAGAACCA AACCCCAAATTC 5011 TAAACTTCCT TACCAAGATT TAGGAACAA CAGAACCA AACCCCAAATTC 5011 TAAACTTCCT TACCAAGATT TAATTGAAA TTTTTTTTTT							
25 4251 TGATTTGCA GARTTATTTG ACAGTTTGCC TATTTCCAGA CAGCCCTGAA 4351 AATCAAATTC TGCTCGATT CAGGGGAAA CATTGCCCTT GGCTACGTCA 4351 AATCAAATTC TGCTCGATT CAGGGGAAA CATTGCCCTT GGCTACGTCA 4451 GCTCTCTCC CGAGCCCAGA TACGGAAGA CATTGGTC CTGCCCGCGG 4401 CTTCCACTTC GTGTATCAAG CTGTTCCTCG TACCAGTGCA CCCAATGCA 4451 GCTCTGTCCC CGAGCCCAGA TACGGAAGAA GAGTGGTC TGGTGTTCTC 4551 TCCACGGCG CTCCACTGCC ATCGTCCAATCCAC CCGGGATACC TGCTTCAGGG 4601 ACGACACGAT CCCCAGCTCT GTGGTACCAC CCGGGATACC TGCTTCAGGG 4601 ACGACACGAT CCCCAGCTCT GTGGTACCAC CCAACGCCTA ACGAGTGGAA 4651 CGAAGAGGTA CACCCAGCTCT GTGGTACCCT GCAGCCCAA TTTCACTCAA 4651 CGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTGAGCCAA TTTCACTCAA 4651 CCAACTGCTA AATACGAAGA TCATACGTTAC GGAGGGGCCCCACAA CAGCACACACACACACACACA							
4351 GCTCACTCT GGGGTCTCAC TCAGGGGAAA CATTGCCCTT GGCTACGTCA 4351 AATCAAATTC TGCTCCGATT CAGTGGAAAGA CATTGCCCTT GGCTACGTCA 4351 CATCCACTTC GGGGTACAAGA CATTGCCCTT GGCTACGTCA 4401 CTTCCACTTC GTGTATCAAG CTGTTCCTG TACCAGTGCA CACCAATGCA 4451 GCTCTGTCCC CGAGCCCAGA TACGGAAGGA GAATTGGTTC TGGTCTCAGTGC 4551 TTCCACGGGG CTCCACTGCC AGTCCGTCCC CAACGCCTTG GCAAGCGTGCA 4601 ACGACACGAT CCCCACTGCT GTGGTACCCT GCAGTGGAACGA CACCAGTGGA 4601 ACGACACGAT CCCCACTGCT GTGGTACCC GCAGTGCAA TACGGAACAA 4701 CTGAACGTA CACCAGCTGT GTGGTACCC GCAGTGCAA TACGGAACAA 4701 CTGAACTGA TATATGGAACAA CCGGGGCAA TACGGAACAA 4701 CTGAACTGA CAGTTTTGCC ACCGGCCTAG CCTGAGCCAT ACGGAACAA 4701 CACGATGGT GGGATTTTGCC ACCGACCGAA ACTGGGACTC CTTGAGACCA 4801 CACGATGGT GGGATTTTGCC ACCGACCAGA ACTGGGACTC CTTGAGATCA 4801 CACGATGGT GGGATGTGAC CGCACCCAGA CTGGGAACT CCTTGAGATC 4801 CACGATGGT GGAATTCAGA ACAGTACTC TACCTCCAT 4901 TCCAGCTGC GCACTGCTGA ACAGTACTC TACCTCCATT 4901 TCCAGTTGGC GATTGGTCA CCAAGAACCA GCCCTCCCCA GCAACAGCAC 4851 ACTGTAGGT CACCTGCAG GCACTCCTG GTTTCCACCT GGAATACAAA 4951 ACTGTAGGT CACCTGCAG GCACTCCTG GTTTCCACCT GGAATACAAA 4951 ACGTTAGGG GATCGGTACA TGGTGAACCA CGTGCTCCC TTCCAGTGCG 5051 AGCCCGGGTA CACCCTGCAG GCCCTCCCC ACATTTCCT TTCCAGTGCG 5101 ACCGTTCCC GTTGGAACTA TCGGTCTCC CTGTGCATT CACCTCAGGGA 5501 GTTCTTACCC CAACAACTTA GACTGCACT GCAGGACTC ATTACCCAGC 5501 GTTCTTACCC CAACAACTTA GACTGCACT GCAGGATCT ATTACCCAGC 5501 GACAATTTAG CGCACAGATT TAGGACCAAC ACGGAACCAG CCCCTCCCAG 5501 TGACTTCCTT GAAATACAAA ATGGAACCTA CCACCAGC CCCATGATG 5401 GAAAACCCTCA TCCACTTTTA TAGGAACAAA ACGGAACCG AAACCGAAT 5401 GAAAACCCTCA TCCACTTTTA TAGGAACAAA ACGGAACCG AAACCGAAT 5401 GAAAACCCTCA TCCACATTTTA TAGGAACAAA ACGGAACCG AAACCGAAT 5401 GAAAACCCCAATTTTA TAGGAACAAA ACGGAACCG GGCAGCGAT 5401 CACACAGGAC CACCAGGGG TTCGCC TTGTTTTTTTTTT	•						
4301 GCTCACTCTC GGGGTCTCAC TCAGGGGAAA CATTGCCCTT GGCTACGTCA 4451 AATCAAATTC TGCTCCGATT CAGTGCAAGA AGCGGTGCCT CTGCCCGCGG 4451 GCTCGCTCC CGAGCCCAGA TACGGAAGGA GAATTGGTTC TGAGTTTCT 30 4501 GCCGGCTCCA TCGTCCGATT CGAGTGCAAGA GAATTGGTTC TGAGTTTTCT 4501 ACGACACGAT CCCCACTGCC AGATCGCAAGGA GAATTGGTTC TGAGTTTTCT 4501 CGAAGAGGAT CCCCAGCTCT GTGGTACCAC CGGGAATACC TGCTTCAGGGA 4601 ACGACACGAT CCCCAGCTCT GTGGTACCC GAAGGCCTTG GCACAGTGGAA 4701 CTTGAACTCT ATATTGGAACAA TCATAGTTAC GGAGGGCAA TTTCACTCAA 4701 CTTGAACTCT ATATTGGAACAA TCATAGTTAC GGAGGGCAA ACGGAAACAA 4701 CACGATGGTG GGAATTGCAC ACGGACCAGA ACTGGGAACC CCTTGAGACCA 4801 CACGATGGTG GGAATGCTGC ACGGACCAGA ACTGGGAACC CCTTGAGATC 4801 CACGATGCTG GGCATGTCAC ACGCACCAGA CTGGGAACCA CTCCCCAGA CTCCCAGACACAACAA 4951 ACTGTAGGTC CACACACACAC CCCACCAGA CTGGGAACCA GCCCTCCCCA GCACCACACA 4951 ACCGTCGCA CATTAGTGTG GCACCCAGA CTGGGAACCA GCCCTCCCCA GCACCACACA 4951 ACCGTCGCC GTTGGTACC CCAAGAACCA GCCCTCCCCC GCAACAACAA 4951 ACCGTTGCC GTTGGAACTA TCGTGCAGCC GCTCCCCC GCAACAACAA 4951 ACCGTTCGC GTTGGAACTA TCGTGCACC GCTCCCCC GCAACCAGCA 5001 ACCGTTCCC GTTGGAACTA TCCGTCTCC CTTCCAGTGCG 5011 ACCGTTCCC CAACAACTTA GACTGCACCT GGAGGACCA CTGCGAGCCC GGCTTCCCC 5011 ACCGTTCCC CAACAACTTA GACTGCACCT GGAGGACCA CTCCCCCA GCACCAGCA 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCT CTCCAGTGCG 5201 GTCTCTCCC CAACAACTTA GACTGCACCT GGAGGATCT ATTCCCAGC 5201 GTATGTGC CACATATTCA GTTTCTGAAT TTTCTTACCC AGCCTCCCAG 5201 TCACTTCCT CAACAACTTA GACTGCACCT GGAGGATCT ATTACCAATC 5301 TCACTTCCT TACCACGTGC CCCCCGCGG CCCTGCTGAG CACAACACATT 5401 GAAACCTCAA TCCACTTTTA TAGTGACCAT TCGCAACACC GGCATAATCA 5501 TCACTTCCT TACCACGTG TTCCCCCCCGCG CCCCTGTGGA CACAACACAC 5501 CTAAATACAC TTCTTACATG TAAATTGTAT TTTTTGGTGG AACCCAAAC 5601 TCTGGTTAGT ACTGGAACAC TTGGCTTTT TAGTGAACAC ACCAGGCC CCCATGATTG 5501 CTAAATACAC TTCCTACCAGC CTTTTTA TAGTGACAA AGCGGAACACA ACCAGGACCA TTCCACCTAACC GGCAACAACACA ACCAGGACACAACA ACCAGGACCA TTCCACGCC TTCCCCCAGGCG CCCACGAGCC CACAACAACACA ACCAGGACACA ACCAGGACACA TTCCACTTTT TAGTGGAACAA ACGGGAACACA ACCAGGCC AACAACACACA TTCCACTTTT TAGTGGAACAAC	25						
4351 AATCAAATTC GGTTATCAAG 4401 CTTCACTTC GTGTATCAAG 4451 GCTCTGTCCC CGAGCCCAGA TACGGAAGGA GAATTGGTTC TGAGTTTCTT 30 4501 GCCGGCTCCA TCGTCCCAGTT CGAGTGCAC CCGGGATACC TGCTTCAGGG 4551 TTCCACGGGG CTCCACTGCC AGTCCGTGCC CAAGGCCTTG GCAGTGGCA 4601 ACGACACCAGT CCCCAGCTGT GTGGTACCCT GCACTGGCAA TTTCACTCAGG 4601 ACGACACCAGT CCCCAGCTGT GTGGTACCCT GCACTGGCAA TTTCACTCAA 4651 CGAAGAGGTA CAATCCTGTC CCCGGCTAC CCTGAGCCAT ACGGAAACAA 4701 CTTGAACTGT ATATGGAACA TCATAGTTAC GGAGGGCTC GGAATACAA 4801 CACGATGGTG GGGATGTGAC CGCACCAGA CTGGGAACT CCTTCAGGAT 4801 CACGATGCG GCACTGCTA ACGAGACCA ACGGAACCA ACGAAACAA 4901 TCCAGTCTGA CGACGCACCAGA CTGGGAACT TCTCAGGAC 4851 ACCGGTCCAC GCACTGCTA ACAGTACTT CAACCAACT TACCTGCATT 4901 TCCAGTCTGA CATAGTGTG GCACCTGCTG GCACTCCCCA GCAACAGCAT 4951 ACTGTAGGTC TTGCTGCAT CCAAGAACCA GCCCTCCCCA GCAACAGCAT 4951 ACCGGGGTA CACCTGCAG GCCGTCTCC CACCAACCACT TACCCAGTG 5001 ACCGGTCCC GTTGGAACTA TCGGTCACC CTTCCAGTGCG 5151 AGCGCGGGTA CACCACTCGAG GGCCGTTCCC CATTCCCAGGG 5151 AGCGCGGTA CACCACCTGG GGCGTTCCC CACTTCCCG GGCTTCCCAGGG 5151 AGCCCGGGTA CACCACTCTG GGGTGTGCT CCTCACTGCC GGCTTCCCAG 5201 GTTCTACCC CAACAACTTA GCTCCCCC GGAGTCTC CACACCTGTG 5301 TGACTTCCT GAAATTCAA ATGGACCAT CCTCCACCAG CCCTGCTGCAG 5301 TGACTTCCTT GAAATTCAA ATGGACCAT TCCCTCTGAGCCC GGCTTCCCAG 5301 TGACTTCCTT GAAATTCAAA ATGGACCAT TCCCTGTGAG CCCTGCTGAAC 5301 TGACTTCCTT TACCACACTTA TACTGCACACCAG CCCTGCTGAGACCA ACACCATTA 5301 TGACTTCCTT TACCACACTTA TACCACACCAG CCCTGCTGCAG CACAACCAT 5301 TGACTTCCTT TACCACACTTA TACCACACCAG CCCTGCTGCAG CACAACCAT 5301 TGACTTCCTT TACCACACTTA TACCACACCAG CCCTGCTGCAG CACAACCAT 5301 TGACTTCCTT TACCACAGACCA TTCCTGTAGC CACACCAGC CCCTGCTGCAG CACACCAGC CCCTGCTGCAG CACACACCAG CCCTGCTGCAG CACACCAGC CCCTGCTGCAG CACACCAG CCCTGCTGCAG CACACACCAG CCCTGCTGCAG CACACCAGACCA TTCTTGAAT TTAATTCTAC GACGAGCCA CCCGGGAATAA TTCTCCCCCGG CCGGCGGAGCG GCGGATACCG CACACCAGCACC CACACACCAG CCCTGCTGCAG CACCCAGGAG TTTTATTGT TAAAATATA TTCGCCGCCG GCGC	25						
4401 CTTCCACTTC GTGTATCAG CTGTTTCTCG TACCAGTGAC ACCCAATGCA 4451 GCTCTGTCCC CGAGCCCAGA TACGGAAGGA GAATTGGTTT TGAGTTTTCT 4501 GCCGGCTCCA TCGTCCGATT CGAGTGCACC CCGGGATACC TGCTTCAGGG 4551 TTCCACGGCG CTCCACTGC AGTCCGTCC CAACGCCTTG GCACAGTGGA 4601 ACGACACGAT CCCCAGCTGT GTGTTACCT GCAGTGGCAA TTTCACTCAA 4601 CGAGACAGAT CCCCAGCTGT GTGTTACCT GCAGTGGCAA TTTCACTCAA 4601 CGAGACAGAT CAATCCTGT CCCCGGCTAC CCTGAGCCAT TTTCACTCAA 4701 CTTGAACTGT ATATGGAAGA TCATGATTAC GGAGGGCTCG GCAATCCAGA 4801 CACCAGTGGT CAGTTTTGCC ACGAGCAGA ACTGGGAACCA CTTGAGACA 4801 CACCAGTGGA CAGTTTTGCC ACGAGCAGA ACTGGGAACCT CCTTGAGATC 4801 CACCAGTGGA CGACTCGTGA CACCCAGA CTGGGAACCT TCCTGAGCAC 4851 CACAGTACCG GCACTGCTGA ACACTACTTC CAACCAACTC TACCTGCAT 4901 TCCAGTCTGA CATTAGTGTG GCAGCTGTG GTTTCCACCT GAATACAAA 4901 TCCAGTCTGA CATTAGTGTG GCAGCTGTG GTTTCCACCT GAATACAAA 4901 ACCGTTGCG GTTGGAACACA GCCCTCCCCA GCAACAGCAT 4951 ACGCTGAGACA CACCTTCGAG GGCCGTTCCC TTCCAGTGG 5001 CAAAATCGGA GATCGGTACA TGGTGAACCA GCCCTCCCCA GCAACAGCAT 5001 ACCGTTCGCC GTTGGAACTA TCCGTCTCCC CTGTGCATTG CAACCAGGCA 5101 ACCGTTCCCC CAACAACTTA GACCTCTCCC CTGTGCATTG CAACCTGTGG 5151 AGGCACGGTA CACCTTTCG GTGTGTCCC CTGTGCATTG CAACCTGTGG 5151 AGGCACCTC TCCTCCAG GGCCGTTCCC CTGTGCATTG CAACCTGTGG 5151 GACAATTTAG CGCACCTTG GTTTCTCACC GAGGATCTC ATTACCCAGC 5201 GTTCTTACCC CAACAACTTA GACTCCACCC GGAGGATCT ATTACCCAGC 5201 GTACTTCCTT GAAATTCAA ATGGACCTT CCACCAGC CCCATGATTG 5301 TGACTTCCTT GAAATTCAA ATGGACCTT CCACCAGC CCCATGATTG 5301 TGACTTCCTT TACCAAGTTA TAGGACAACA ACGAGAACCA ACCCAGCAT 5401 GAAACCCCAA TTCCACTTTA TAGTGACCAT CACCACACCA		4301	GCTCACTCTC	GGGGTCTCAC	TCAGGGGAAA	CATTGCCCTT	GGCTACGTCA
451 GCCGGCTCCA TCGTCCGATT CGACTGCAAC CCGGGATACC TGCTTCAGGG 4551 TTCCACGGGG CTCCACTGCC AGTCCGTGCC CAACGCCTTG GCACGGATACC TGCTCCAGTGCATT CCACGGGATACC TCGCTCGAGTGCAC CAACGCCTTG GCACGGAGGAACAA 4601 ACGACACGAT CCCCAGCTGT GTGGTACCCT GCAGTGGCAA TTTCACTCAA 4651 CGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTGAGCCAT AGGGAACAA 4701 CTTGAACTGA TATTGGAACA TCATTAGTAC GGAGGGCCTG GGAACAACAA 4701 CTTGAACTGAT CAGTTTTGCC ACGGACAGA ACTGGGACTC GGAACAACAA 4701 CACGATAGGT CAGTTTTGCC ACGGACAGA ACTGGGACTC CCTTGAGATC AGGAACAA 4851 CAACGATCGG GCACTGCTGA ACAGTACTT CAACCAACTC TTCCAGTAC AGGAACAA 4901 TCCAGTTCGA CACTTGCTG ACAGTACTT CAACCAACTC TACCTGCATT 4901 TCCAGTTCGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GCAACAACAA 4951 ACTGTAGGTC TTGCTGCAT CCAACAACACA GCCTCCCCA GCAACAACAA 4951 ACTGTAGGTC TTGCTGCAT CCAACAACCA GCCTCCCCA GCAACAGCAT AGGCAGACAA AGGCAACAACAA 4951 ACCGTTCGCA GACCGTGCGA GCCGTTCCC ACATTTCCTG TACCAGGGA 5001 ACCGTTGCG GTTGGATA TGCGTGTCC CACATTTCCTG TACCCAGGG 5101 ACCGTTGCC GACCACTTG GGCGGTTCC ACATTTCCTG TACCCAGGG 5101 ACCGTTGCC GACCACTTG GGCGGTTCC ACATTTCCTG TATCCCAGGG 5101 ACCGTTCCC CAACAACTTA GACTGCACC CCTGTGCATTG CAACCTGTGG 5151 AGGCACGCTG ACACACTTG GTGTGTGAT CCTGAGCCCC GGCTTCCCAGC 5001 GTCTCTACC CAACAACTTA GACTGCACT CCAGCCCC GGCTTCCCAGCCC GGAGGATCCC CAACAACTTA GACTGCACT TCCAGCCCC GGCTTCCCAGCCCT GGAGGATCTC ATCCCATC AAAATTCAA ATGGACCTTA CCAACACCTTG GGAGGATCCA CCCAGTGTGG CACAACAACTA TCCACACCAGC CCCATGTTG 5351 GACAATTAGA ATCCAACACTTA TACCAACA ACGAGAACCA AAACCCAAT 5401 GAAACCCAAC TCCACCCGCG CCCTGCTGAG CACAACCGAT TCCACCAGC CCCATGTTG CAACAACTA TCCACACCAGC TCCACCAGC CCCATGTTG CAACAACTA TCCACACCAGC CCCATGTTG CAACAACTA TCCACACCAGC CCCATGTTG CAACAACAC TCCACCAGC CCCATGTTG CAACAACAC TCCACACGCA CACAACGCAT TCCACACCAGC CCCATGTTG CAACAACAC ACAACCCATA ACCGAACAC TCCACAGACC TCCACAGACC TCCACAGACC TCCACCAGACC TCCACCAGACC TCCACAGACC TCCACAGACC TCCACAGACC TCCACAGACC TCCACAGACC AAACCCAAACAC TCCACAGACC TCCA		4351	AATCAAATTC	TGCTCCGATT	CAGTGCAAAG	AGCGGTGCCT	CTGCCCGCGG
451 GCCGGCTCCA TCGTCCGATT CGACTGCAAC CCGGGATACC TGCTTCAGGG 4551 TTCCACGGGG CTCCACTGCC AGTCCGTGCC CAACGCCTTG GCACGGATACC TGCTCCAGTGCATT CCACGGGATACC TCGCTCGAGTGCAC CAACGCCTTG GCACGGAGGAACAA 4601 ACGACACGAT CCCCAGCTGT GTGGTACCCT GCAGTGGCAA TTTCACTCAA 4651 CGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTGAGCCAT AGGGAACAA 4701 CTTGAACTGA TATTGGAACA TCATTAGTAC GGAGGGCCTG GGAACAACAA 4701 CTTGAACTGAT CAGTTTTGCC ACGGACAGA ACTGGGACTC GGAACAACAA 4701 CACGATAGGT CAGTTTTGCC ACGGACAGA ACTGGGACTC CCTTGAGATC AGGAACAA 4851 CAACGATCGG GCACTGCTGA ACAGTACTT CAACCAACTC TTCCAGTAC AGGAACAA 4901 TCCAGTTCGA CACTTGCTG ACAGTACTT CAACCAACTC TACCTGCATT 4901 TCCAGTTCGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GCAACAACAA 4951 ACTGTAGGTC TTGCTGCAT CCAACAACACA GCCTCCCCA GCAACAACAA 4951 ACTGTAGGTC TTGCTGCAT CCAACAACCA GCCTCCCCA GCAACAGCAT AGGCAGACAA AGGCAACAACAA 4951 ACCGTTCGCA GACCGTGCGA GCCGTTCCC ACATTTCCTG TACCAGGGA 5001 ACCGTTGCG GTTGGATA TGCGTGTCC CACATTTCCTG TACCCAGGG 5101 ACCGTTGCC GACCACTTG GGCGGTTCC ACATTTCCTG TACCCAGGG 5101 ACCGTTGCC GACCACTTG GGCGGTTCC ACATTTCCTG TATCCCAGGG 5101 ACCGTTCCC CAACAACTTA GACTGCACC CCTGTGCATTG CAACCTGTGG 5151 AGGCACGCTG ACACACTTG GTGTGTGAT CCTGAGCCCC GGCTTCCCAGC 5001 GTCTCTACC CAACAACTTA GACTGCACT CCAGCCCC GGCTTCCCAGCCC GGAGGATCCC CAACAACTTA GACTGCACT TCCAGCCCC GGCTTCCCAGCCCT GGAGGATCTC ATCCCATC AAAATTCAA ATGGACCTTA CCAACACCTTG GGAGGATCCA CCCAGTGTGG CACAACAACTA TCCACACCAGC CCCATGTTG 5351 GACAATTAGA ATCCAACACTTA TACCAACA ACGAGAACCA AAACCCAAT 5401 GAAACCCAAC TCCACCCGCG CCCTGCTGAG CACAACCGAT TCCACCAGC CCCATGTTG CAACAACTA TCCACACCAGC TCCACCAGC CCCATGTTG CAACAACTA TCCACACCAGC CCCATGTTG CAACAACTA TCCACACCAGC CCCATGTTG CAACAACAC TCCACCAGC CCCATGTTG CAACAACAC TCCACACGCA CACAACGCAT TCCACACCAGC CCCATGTTG CAACAACAC ACAACCCATA ACCGAACAC TCCACAGACC TCCACAGACC TCCACAGACC TCCACCAGACC TCCACCAGACC TCCACAGACC TCCACAGACC TCCACAGACC TCCACAGACC TCCACAGACC AAACCCAAACAC TCCACAGACC TCCA		4401	CTTCCACTTC	GTGTATCAAG	CTGTTCCTCG	TACCAGTGAC	ACCCAATGCA
4501 GCCGGCTCCA TCGTCCGATT CGAGTGCAC CCGGGATACC TGCTTCAGGG 4551 TTCCACGGCG CTCCACTGCC, AGTCCGTGCC CAACGCCTTG GCACAGTGGA 4601 ACGACACGAT CCCCAGCTGT GTGGTACCT GCAGTGGCAA TTTCACTCAA 4651 CGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTGAGCCAT ACGGAAACAA 4701 CTTGAACTGT ATATGGAACA TCATAGTTAC GGAGGGCTCG GCAATTCAGA 4801 CACGATGGTG GGGATGTGCA ACGGAGAGACA ACGGAAACAA 4751 TCCAAGTGTG GGAGTGTGAC CGCACCCAGA CTGGGAACCT TCCAGGCAC 4851 CACAGTACCG GCACTGCTGA ACACTACTTC CAACCAACTC TACCTGGATC 4801 TCCAGTTCTGA CATTAGTGTG GCACCTGGGTACT TACCTGGATC 4901 TCCAGTCTGA CATTAGTGTG GCACCTGGGTACT TACCTGCATT 4901 TCCAGTCTGA CATTAGTGTG GCACCTGCTG GTTTCCACCT GCAATACAAA 4951 ACTGTAGGTC TTGCTGCATG CCAACAACCA GCCCTCCCCA GCAATACAAA 4951 ACCGTTAGCA CACCTTGCATG CGCACCTCCCA GCAATACAAA 4951 ACCGTTCGCC GTTGGAACTA TCGTCTCC CTGTGCACCT TACCTGCAGCAC 5001 CAAAATCGGA GATCGGTACA TGGTGAACCA GCCCTCCCCA GCAATCAGCAC 5010 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGTGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGTGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCACCC CGGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTCACCC CGGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GTTCCACCTTG CCAACCAGCAC 5201 GTCTTCCTC CAACAACTTA GTTCCACCTTAC CAACCTGTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG 5451 TAAACTTGCT TACCAAGATA TGGAACACA ACGAAAACC GACAACGCAT 5451 TAAACTTGCT TACCAAGATA TGGAACACA ACGAAAACC GACAACGAT 5451 TAAACTTGCT TACCAAGATA TGGAACACA ACGAGAACC GACAACGCAT 5451 TAAACTTGCT TACCAAGATA TGGAACACA ACGAGAACC GACAAGCCA 5501 CTACAGGAA CACCAGTGGG TTCGCCCTC CTGCCCCA GACAACCGAC 5501 ATTTATCAAG ACGGGAATTG CAATGGAAAA TTAAGTATAA ATTTCCCTAA 5551 CTGGTTCCAA GCTGTACCA GTGGAATAAT TTTTTGGTGG AATCTCCCTAA 5551 CTACGCCGT TTATCCACTTTT TAAAGAAAA AGGGAAAACC GGCAAGAGCC 5701 ATTTATCAG ACCGGGG TTTCTTTTTT TTAAGAAAAC GGCAAGACC 5701 ATTTATCAG ACCGGGG TTTCTTTTTTTTTTTTTTTTT							
4551 TTCCACGGCG CTCCACTGCC AGTCCGTGCC CAACGCCTTG GCACAGTGCA 4601 ACGACACGAT CCCCAGCTGT GTGGTACCCT GCACTGGCAA TTTCACTCAA 4651 CGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTCAGCCAT ACGACACCAA 4701 CTTGAACTGT ATATGGAAGA TCATAGTTAC GGAGGGCCAT ACGAAACCAA 4801 CACGATGGTG GGGATGTAC CGCACCAGA ACTGGGAGCT CCTTGAGATC 4801 CACGATGCG GCACTGCTCA ACAGCACCAGA ACTGGGAGCT TCTCAGGCAC 4851 CACAGTACCG GCACTGCTCA ACAGTACTT CAACCAACTC TACCTGCATT 4901 TCCAGTCTGA CATTAGTGT GCAGCTGCTG GTTTCCACCT GGAATCAAA 4951 ACTGTAGGTC TTGCTGATG CCAGAAACCA GCCCTCCCCA GCAACACACT 5001 CAAAATCGGA GATCGGTACA TGGTGAACCA CCTCCCCA GCAACACACT 5001 ACCCGTCGCC GTTGGAACTA TCGTCTCCC CTCTCCAGTGCG 5051 AGCCCGGGTA CACCCTGCAG GCCCTTCCC ACATTTCCTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG ACCACTTGG GTGTGTGTTC CTAGGCCC GCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GCCTCTCCC GGAGGATCC ACACTGTGG 5201 GTTCTTACCC CAACAACTTA GACTGCACT CCTCAGGCCC GCCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCC AACCTGTGG 5301 TGACTTCCTT GAAATTCAAA ATGGACCTT CCTCAGCCCC CCCTTCCCAG 5301 TGACTTCCTT GAAATTCAAA ATGGACCTT TCCCCCTGGG CCCTGATGTG 5351 GACAATTTAG CGGCACCGAT CTCCCCCGCG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC AACCCAACC 5301 TAAACTTGCT TCCACTTTTA TAGTGACCAT TCGCAAAACC ACCAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC AACCCAAAC 5501 CTAAAATACAC TTCTTACATG TAAATTGTAT TTTTTTGGTG AACCCAAAC 5501 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTTGGTG AACCCAAAC 5501 TCTGGTTACGA GCTTGTACGA GTGGAATAAT TTTTTTTGGTG AACCAACCAAC 5501 ACTGGGAAC CACCAGTGGG TTCGCCTTTT TCAAGAACAA ACGAACCAAACAAACA ACGAACCAAAT 5601 TCTGGTTACGA GCTGGAACCA TTCTTTTTT TCAAACTAGA ACCAAGACCAAACAA ACGAAGACCAAAT 5601 TCTGGGGAACC TCTCTCCCCCGCG GCCCGGGGCG GCGCGGGGCG 5701 ACTTCACGCCT TAATCCCACC ACTTGGAAAA ACGAGAACCA ACCAAGCCA 5701 ACTTCACGCCT TAATCCCACA CCTTTTTATTGTG ACCCAATCCA GCCCGCGGGCG 5701 ACTTCACGCCT TAATCCCACA TCTTGGAAC CCCGGGGGG GCCCGGGGCG 5701 ACTTCACGCCT TAATCCCAC ACTTGGAAAC CCCGGGGGG GCCCGGGGGG 5701 AGGTCACCG GAGCCCTG TCTTTTTTTTATTGTG ACCCACAACC CCGGGGGCG 570	30						
4601 ACGACAGGAT CCCCAGCTGT GTGGTACCCT GCAGGCCAT ACGAGAACAA 4651 CGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTGAGCCAT ACGAAACAA 4701 CTTGAACTGT ATATGGAAGA TCATAGTTAC GGAGGGCTCG GCAATCAGAA 4751 TCCAAGTGAT CAGTTTTGCC ACGGAGCAGA ACTGGGACTC CCTTGAGATC 4801 CACGATGGTG GGGATGTGAC CGCACCCAGA CTGGGAAGCT TCTCAGGCACT 4801 TCCAGTCTGA CATTAGTGTA CAGCAGCACC CTGGGAAGCT TCTCAGGCAC 4851 CACAGTACCG GCACTGCTGA ACAGTACTTC CAACCACTC TACCTGCATT 4901 TCCAGTCTGA CATTAGTGTA GCAGCTGCTG GTTTCCACCT GCAATCACAA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGCAT 4951 ACCGTTCGCA GATCGGTACA TGGTGAACGA GCCCTCCCCA GCAACAGCAT 5001 CAAAATCGGA GATCGGTACA TGGTGAACGA CCTGCTCC TTCCAGTGCG 5051 ACCCGTGCGA GCCCTTCCC CTGTGCATTC CTACCTGCGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGCGG 5101 ACCGTTCCCC CAACAACTTA GCCTCACCC CTGTGCATTG CAACCTGCGG 5101 ACCGTTCCCC CAACAACTTA GACTCACCC GGAGTACCA ATTACCCATC 5201 GTTCTTACCC CAACAACTTA GACTCACCT GCAGGATCT ATTACCCATC 5201 GTTCTTACCC CAACAACTTA GACTCACCT GCAGGATCT ATTACCCATC 5201 GAAACCCTCA TCCACATTTA ATGGACCACT CCCACACACC CCCATGATTG 5351 GACAATTTAG GGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACCCAT 5401 GAAACCCCA TCCACTTTTA TAGGAACACA CGAGAACC ACAACCCAT 5401 GAAACCCTCA TCCACTTTTA TAGGAACAAAC GGCAAAACC ACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGGAACAAAC ACGAGAACC GACAACCCAAC 5501 CTAAATACAC TTCTTACATG TAAATTGTAT TTTTTGGTGG AATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTTGGTGG AATCCCATA 5551 CTGGTTCAA GCTGTGACGA GTGGAATAAT TTTTTTGGTGG AATCTCCCTAA 5551 CTGGTTCAA GCGGAACCG TTCTGCTTTT TAAAATAACA GGCAGACCG AACCCAAAC 5551 CAGACGGAAC CACCAGTGGG TTTTTTTTTTTTTTTTT							
4651 CGGAAGAGGTA CAATCCTGTC CCCCGGCTAC CCTGAGCCAT ACGGAAACAA 4701 CTTGAACTGT ATATGGAACA TCATAGTTAC GAGGGCTCG GGAATTCAGA 4751 TCCAAGTGAT CAGTTTTGCC ACGGAGCAGA ACTGGGACTC CCTTGAGATC 4801 CACGATGGTG GGGATGTGAC CGCACCCAGA CTGGGAACCT TCTCAGGCAC 4851 CACAGTACCG GCACTGCTGA ACAGTACTTC CAACCAACCT TACCTGCATT 4901 TCCAGTCTGA CATTAGTTG GCAGCTGCTG GTTTCCACCT GGAATACAAA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGCAT 5001 CAAAATCGGA GATCGGTACA TGGTGAACCA GCCCTCCCCA GCAACAGCAT 5001 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGTGG 5151 AGGACGCTG ACACCTTGG GTGGTGTAT CCTGAGCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCT ATTACCCATC 45 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACACCCATC 5351 GACAATTTAG CGGCACGGAT CTCCCCCGCG CCCTGCTGAG CACACCCATC 5301 TGACTTCCTT TACCAAGNTA TGGAACAACA ACGAGAACCG ACACCCAAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTG AATGTTGGTT 5601 CTAAATACAC TTCTTCACTG TAAATTCTAT TTTTTTGGTG AATGTTGGTT 5601 TCTGGTTTAGT AGTGGAACAC TTGTTTTTT TGGAACAACA GACGAACCG TACGGAACCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAAAAATTA TTGGCAGCAC GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAAAAATTA TTGGCAGCCC GGCGGGTGG 5851 CACACCGCGG TTAATCCCAGC ACTTTGGAACAC GGCCAGGCCG GCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAACAC CACCAGTGGG GCGGAGCCG GCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAACAC CACCAGTGGG GCGGAGCGG GCGGATCACG 5861 CACTTCAGGGA ATCAAGATTTA GCCGGCAAA ACCGAGGCCG GCGGGTGGG 5851 ACAAAAATTA CCAGAAAATTA GCCGGCAATAG GCCGGAGCCG GCGGGTGGG 6051 AGCTTGCAGT GAGAGAGAT CCCGGCACTG CCCTGTGACC CGGGGGGGGG 6051 AGCTTGCAGT GAGAGAGAT CCCGGCACTG CCCTGTGACC CGGGGGGCGG 6051 AGCTTGCAGT GAGAGAGAT CCCGGCACTG CCCTGTGACC CGGGGGCGG 6051 AGCTTGCAGT GAGAGAGAT CCCCGCCACT							
4701 CTTGAACTGT ATATGGAAGA TCATAGTTAC GGAGGGCTCG GGAATTCAGA 4751 TCCAAGTGAT CAGTTTTGCC ACGAGCAGA ACTGGGACTC CCTTGAGATC 4801 CACGATGCTG GGGATGTGAC CGCACCCAGA CTGGGAACTC TCTCAGGCAC 4851 CACAGTACCG GCACTGCTGA ACAGTACTT CAACCAACTC TACCTGCATT 4901 TCCAGTCTGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GGAATACAAA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCTCCCCA GCAACAGCAT TTGCTGCATG CACCCTGCAG GGCCGTTCC ACATTTCCTG TACCTGCAGC 5001 CAAAATCGGA GATCGGTACA TGGTGAACCA CGTGCTCTCC TTCCAGTGCG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGTACT CTGAGCCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATACCCATC 45 5251 GGCTATGGTG CACATATTCA GACTGCACCT GGAGGATCTC ATACCCATC 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5301 GACAATTTAG 5301 GACAATTTAG 5301 GACAATTTAG 5301 GACAATTTAG 5301 GACAATTTAG 5301 GACAATTTAG 5301 TAAACCTCA TCCACTTTTA TAGGGACCAAC CCCCTGCTGAC CACAACCATC 5401 GAAACCCTCA TCCACCTTTTA TAGGTGACCAT TCGCAAAACC GCCAGAGCCT 5401 GAAACCCTCA TCCACTTTTA TAGGTGACCAT TCGCAAAACC GACAACGCAT 5501 CTAAATACAC TCCACTTTTA TAGGTGACCAT TCGCAAAACC GACAACCAAT 5501 CTAAATACAC TCCACTTTTA TAGGTGACCAT TTTTTGGCAGA AACCCAAAT 5501 CTGGTTACT AGGGGAACC TTGTTGTAT TTTTTTGGCAGAGCC 5701 ATTTATCAAG ACGGGAATG TTGGCAATAAT TTTTTTGGCAGAACC 5701 ATTTATCAAG ACGGGAATG TTGGCATTAT TTGAAAACACA AGGAGACCG 5701 ATTTATCAAG ACGGGAATG CAATGGGAAA AGAGAAAATA ACGCAGAGCCG 5701 ATTTATCAAG ACGGGAATG CAATGGGAAAATT TTGAAACACA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGG TTTTTTTTTTTTTTTTTT							
4751 TCCAAGTGAT CAGTTTTGCC ACGGAGCAGA ACTGGGACTC CCTTGAGATC 4801 CACCATGGTG GGGATGTACC CGCACCCAGA CTGGGAAGCT TCTCAGGCAC 4851 CACAGTACCG GCACTGCTA ACAGTACTTC CAACCAACTC TACCTGCATT 4901 TCCAGTCTGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GGAATACAAA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGAT 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGACT 5001 CAAAATCGGA GATCGGTACA TGGTGAACGA CGTGCTCTCC TTCCAGTGCG 5051 AGCCCGGGTA CACCTGCAG GGCCGTTCCC CACTTTCCTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGTGAT CCTGAGCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 5301 TGACTTCCTT GAAATTCAA ATGGACCAT TCCACCAGC CCCATGATTG 5301 TGACTTCCTT GAAATTCAAA ATGGACCTAT CCCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTAT TAGTGACCAT TCGCAAAACC GGCAAGGATT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5401 GAAACCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5551 CTAAATACAC TCCTACAGGA TTTAATTGTAT TTAAGTATAA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACACA TTGTTGTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGACCG 5701 ATTATCAAG ACGGAAATT CAATTGTAT TTAAGTAATT AATCCCCTAA 5551 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGACCG 5701 ATTATCAAG ACGGAATT CAATGGAAA AGAGTAATT CAGCAGACCG 5701 ATTATCAAG ACGGAATT CAATGGAAA AGAGTAATT CAGCAGACCG 5701 ATTATCAAG ACGGGAATT CAATGGAAA AGAGTAATT CAGCAGACCG 5701 ATTATCAAG ACGGGAATT CAATGGAAA AGAGTAATT CAGCAGACCA 5651 CAGACGGAG TTTATTTTT TAAGATAATT TTGGCGGCCG GCGCGGTGG 5701 ATTATCAAGA ACGGAATT CAATGGAAA AGAGTAATT CAGCAGACCA 5701 ATTATCAAG ACGGAATT CAATGGAAA ACGGAGACA GCGGAGCGG GCGCGGTGG 5701 ATTATCAAG ACGGAATT CAATGGAAA ACGGAGACA ACTTGGAGACA ACTTGAATTC ACGCAGACCA 5701 AGTTCAGGGA TTCAACTTT TAAAGATAAT TTGGCGGCCG GCGGGTGG 5701 AGTTCAGGACCA TCCTGGCTAA CACGGGGCG GCGGGCGG GCGGGTGG 5701 AGTTCAGGACA ACTTCTGGC							
4801 CACGATGGTG GGGATGTGAC CGCACCCAGA CTGGGAAGCT TCTCAGGCAC 4851 CACAGTACCG GCACTGCTG ACAGTACTT CAACCAACTC TACCTGCATT 4901 TCCAGTCTGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GGAATACAAA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGCAT 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGCAT 5001 CAAAATCGGA GATCGGTACA TGGTGAACGA CGTGCTCTC TTCCAGTGCG 5051 AGCCCGGGTA CACCCTGCAG GGCCGTTCCC ACATTTCCTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGTAT CCTGAGCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 45 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TCTTTCACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCTCA TCCACTTTTA TAGTGACCAT TCGCAAAAAC GGCAAGACCAT 5451 TAAACTTGCT TACCAAGANTA TAGTGACCAT TCGCAAAAAC GGCAAGGCAT 5451 TAAACTTGCT TACCAAGANTA TAGTGACCAT TCGCAAAACC GGCAAGGCAT 5451 TAAACTTGCT TACCAAGANTA TGGAACACA ACGAGAACCG AAACCCAAAT 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATCCCTAA 5551 CTGGTTAGTA AGTGGAACAC TTGTTGTTTT TGAAAAACAGA GGTAAAGCAC 5651 CACAGGGAC CACCAGTGG TTCGCCTTTT CTGCTGCCCA GACAGAGCC 5701 ATTTATCAAG ACGGGAATTG TAAATTGTAT TTAAGTATAA ATCTCCCTAA 5651 CACAGGGAC CACCAGTGG TTCTGCCTTTT TGAAAACAGA GGTAAAGCAC 5651 CACAGGGAC TTCAGGTTTT TAAAGATAAT TTGGCGCCC GGCGGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAAG ACGGGAATTG TAAAGATAAT TTGGCCGCC GGCGGATCACG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCCGCC GGCGGATCACG 5801 CATTCAGGGA ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCTTCTC 5951 ACTAAAAATA CCAAAAATTA CCAGAGAAAC CACGGTGAAAC CCCGTTCTC 5951 ACTAAAAATA CCAAAAATTA CCAGAGAAAC CACGGTGAAAC CCCGTTCTC 5951 ACTAAAAATA CCAAAAATTA CCGGGCATA GCGGGCGG GCGGGATCACG 6051 AGCTTGCAG GAGAGCACA CCCGCCACTG GGCGGCGG CCTGTAGCGCG 6051 AGCTTACCG GAGAGCACAC TCCTGCCACTG CGCGCGCG CCTGTAGCGCG 6051 AGCTTGCAG GCAGGAAAAC CACGGTGAACC CTGGGGGGGGGG	25						
4851 CACAGTACCG GCACTGCTGA ACAGTACTTC CAACCAACTC TACCTGCATT 4901 TCCAGTCTGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GGAATACAAA 4951 ACTGTAGGTC TTGCTGCATG CCCAAGAACCA GCCCTCCCCA GCAACACACA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACACACA 5001 CAAAATCGGA GATCGGTACA TGGTGAACGA CGCCTCCCC TTCCAGTGCG 5051 AGCCCGGGTA CACCCTGCAG GGCCGTTCCC ACATTTCCTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGTAT CCTGAGCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACT GGAGGATCTC ATTACCCATC 45 5251 GGCTATGGTG CACATATTCA GTTTCTAGAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACCCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACACA ACGAGAACCC GACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACACA ACGAGAACCG GAACGCAT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5451 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAAG ACGGGAATTG TAAATTTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCCGCC GGCGGGCG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCCGCC GGCGGATCACG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCCGCC GGCGGATCACG 5801 CATTCAGGAG ATCGGAACCA TCCTGGCTAA CACGGTGGAA CCCGGTTCCC 5951 AGATGAGAAC TCCAGAGCC ACTTTGGAAC CCCGGAGCCG GCGGATCACG 5901 AGGTCAGGAG TTCAAAATTA CCCGCCATG GCGGGCG CCTGTAGTCC 5951 ACTAAAAATA CCAAAAATTA CCCGGCCATG GCGGGCG CCTGTAGTCC 5951 ACTAAAAATA CCAAAAATTA CCGGGCATA GCGGGCGG CCTGTAGCG 6051 AGCTTGCAG GAGGACGAGAT CGCGCCACTG GGCGGGCG CCTGTAGCG 6051 AGCTTGCAG GAGGAGAT CGCGCCACTG CGCGCGCG CCTGTAGCGCG 6051 AGCTTGCAG GAGGAGAT CGCGCCACTG CACTCCAGCC TGGGCGG 6051 AGCTTGCAG GAGGAGAT CGCGCCACTG CA	33				•		
4901 TCCAGTCTGA CATTAGTGTG GCAGCTGCTG GTTTCCACCT GGAATACAAA 4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGCAT 5001 CAAAAATCGGA GATCGGTACA TGGTGAACGA CGTGCTCTCC TTCCAGTGGG 5051 AGCCCGGGTA CACCCTGCAG GGCCGTTCCC ACATTTCCTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGTGAT CCTGAGCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 45 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCAACCAGC CCCTGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GCACAAGCAT 5451 TAAACTTGCT TACCAAGANTA TGGAACACA ACGAGAACCG CACAACGCAT 5451 TAAACTTGCT TACCAAGANTA TGGAACACA ACGAGAACC GACAAGCAT 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTTGGTGG AATGTTGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTTTT TGAAAAACAG AGGTAAACGC 5701 ATTTATCAAG ACGGGAATTC CAACGAGAC CTCCCCAGAGCCG 5701 ATTTATCAAG ACGGGAATTC CAATGGAGAA AGAGTAATTC AGCCAGAGCCG 5701 AGTTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GCGGAGCCG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GCGGAGCCG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GCGGAGCCG 5801 CACTACTCC GCAGACCA TCCTGGCTAA CACGGTGAACC 5901 AGCTCAGGCG GCGAGCCGA CCCTTGGCTAA 6001 CAGCTACTCC GGAGGCTGAG GCAGGANAGT GGCGGGGCG CCTGTAGTCC 6001 CAGCTACTCC GGAGGCTGAG GCAGGANAGT GGCGGGGCG CCTGTAGTCC 6001 CAGCTACTCC GGAGGCTGAG GCAGGANAGT GGCGGGGCG CCTGTAGTCC 6001 CAGCTACTCC GGAGGCGG GCGGCACCT GCGGCGACCG CCTGTAGTCC 6001 CAGCTACTCC GGAGGCGG GCGGCACCT CCCGGCACCT GCCGGCGCG CCTGTAGCC 6001 CAGCTACTCC GGAGGCGAG CCCACCACTC CCCTGCACC CCTGTAGTCC 6001 CAGCTACTCC GGAGGCGAG CCCACCACTC CCCCCCACTC CCCCCACCCCCCCCCC			CACGATGGTG	GGGATGTGAC	CGCACCCAGA	CTGGGAAGCT	TCTCAGGCAC
4951 ACTGTAGGTC TTGCTGCATG CCAAGAACCA GCCCTCCCCA GCAACAGCAT 5001 CAAAATCGGA GATCGGTACA TGGTGAACGA CGTGCTCTCC TTCCAGTGCG 5051 AGCCCGGGTA CACCCTGCAG GGCCGTTCCC CTGTGCATTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGTAT CCTGCAGCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 45 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACACGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACACTT 5401 GAAACCCTCA TCCACTTTAT TAGTGACCAT TCGCAAAACC GGCAAGAGTT 5451 TAAACTTGCT TACCAAGANTA TGGAACAACA ACGAGAACCG AAACCCAAA 5451 TAAACTTGCT TACCAAGANTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5451 TAAACTTGCT TACCAAGANTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTCTTTTTT TGAAAACAGA GGTAAGGAC 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCCACAGAGCCG 5701 ATTTATCAAG ACGGGAACTG CACCAGTGGG TTCGCCTTTT CTGCCCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCCG 5701 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATCCAGCGT TAATCCCAG ACTTTGGAAA AGAGTAATTC ACGCAGAGCCG 5801 CATCCAGCGT TAATCCCAG ACTTTGGAAA AGAGTAATTC ACGCAGAGCCG 5801 CATCCAGGAA ACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATCAAGAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACA TCCTGGCTAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG CCTTGTAGTCC 60 6001 CAGCTACTCG GGAGGCTGAG GCCGGCATA GTGGCGGACCC 60 6001 CAGCTACTCG GAGGACGAG GCGGCATA GTGGCGGACG 60 6001 CAGCTACTCG GAGGACGAT CCGCGCCACTG CACCCGCCC TGGGGCGG 60 60 60 60 60 60 60 60 60 60 60 60 60 6		4851	CACAGTACCG	GCACTGCTGA	ACAGTACTTC	CAACCAACTC	TACCTGCATT
40 5001 CAAAATCGGA GATCGGTACA TGGTGAACGA CGTGCTCTCC TTCCAGTGCG 5051 AGCCCGGGTA CACCCTGCAG GGCCGTTCCC ACATTTCCTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGAT CCTGAGCCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 45 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACC AAACCCAAAT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTAGTATA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGTA ACGAGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAGG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAGG ATCAAGTTTT TAAAGATAAT TTGGCGGCCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCCG GCGGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG CCCGAGGCCG GCGGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG CCCGGGGAGCCG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCCGTCTCT 5951 ACTAAAAATA CCAAAAAATTA GCCGGGCAAA CACGGTGAAC CCCGTCTCT 5951 ACTAAAAATA CCAAAAAATTA GCCGGGCCATA GTGGCGGCCG CCTGTAGTCC 60 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGGAGCGG 6051 AGCTTGCAGT GAGGGAGAAT CGCGCCACTG CACTCCAGC TGGGGGCGG 6051 AGCTTGCAGT GAGGGAGAAT CGCGGCCACTG CACTCCAGC TGGGGGCGG 6051 AGCTTGCAGT GAGGAGAAT CGCGGCCACTG CACTCCAGCC TGGGGGGGGGG		4901	TCCAGTCTGA	CATTAGTGTG	GCAGCTGCTG	GTTTCCACCT	GGAATACAAA
40 5001 CAAAATCGGA GATCGGTACA TGGTGAACGA CGTGCTCTCC TTCCAGTGCG 5051 AGCCCGGGTA CACCCTGCAG GGCCGTTCCC ACATTTCCTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGAT CCTGAGCCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 45 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACC AAACCCAAAT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTAGTATA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGTA ACGAGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAGG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAGG ATCAAGTTTT TAAAGATAAT TTGGCGGCCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCCG GCGGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG CCCGAGGCCG GCGGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG CCCGGGGAGCCG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCCGTCTCT 5951 ACTAAAAATA CCAAAAAATTA GCCGGGCAAA CACGGTGAAC CCCGTCTCT 5951 ACTAAAAATA CCAAAAAATTA GCCGGGCCATA GTGGCGGCCG CCTGTAGTCC 60 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGGAGCGG 6051 AGCTTGCAGT GAGGGAGAAT CGCGCCACTG CACTCCAGC TGGGGGCGG 6051 AGCTTGCAGT GAGGGAGAAT CGCGGCCACTG CACTCCAGC TGGGGGCGG 6051 AGCTTGCAGT GAGGAGAAT CGCGGCCACTG CACTCCAGCC TGGGGGGGGGG		4'951	ACTGTAGGTC	TTGCTGCATG	CCAAGAACCA	GCCCTCCCCA	GCAACAGCAT
5051 AGCCCGGGTA CACCCTGCAG GGCCGTTCCC ACATTCCTG TATGCCAGGG 5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGTAT CCCTGAGCCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 45 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACACGCAT 5401 GAAACCCTCA TCCACTTTA TAGTGACCAT TCGCAAAACC GCCAAGGAT 5401 GAAACCCTCA TCCCACTTTA TAGTGACCAT TCGCAAAACC GGCAAGGAT 5451 TAAACTTGCT TACCAAGANTA TGGAACACA ACGAGAACCG AAACCCAAAT 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGG TTCGCCTTTT TGAAAACAGA GGTAAGGCC 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA ACTCAATTCA GTCCCCCAG 5801 CATTCAGGGA TTCAAGTTT TAAAGAATAAT TTGGCGGCCG GGCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GGCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGGTGG 5851 AGTAAAAATA CCCAAAAATA GCCGGGCATA GTGGCGGCG GCGGGTGCG 5901 AGGTCAGGG GCGGGTGAG GCAGGCATA GTGGCGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGCAACG CACCCAGCC CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGCAACG CACCCAGCCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGCAACG CACCCAGCCG CCCGGGCGG GCGGGGGGGG	40						
5101 ACCGTTCGCC GTTGGAACTA TCCGTCTCCC CTGTGCATTG CAACCTGTGG 5151 AGGGACGCTG AGCACCTTGG GTGGTGTGAT 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5401 GAAACCTCA TCCACTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5501 CTAAATACAC TTCTTACATG TAAATTGTAT TTAAGTATAA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT TGAAAACAGA GGTAAGGACA 5651 AGATTGTAGG AAGACCGGAG TTTTATTTGGT ACCCAAGACCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCCG 5701 ACTTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCCGCCG GCCGCGGGGCG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCCG GCGGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG CCCGAGGCCG GCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG CCCGAGGCGG GCGGGTGG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG CCTGTAGTCC 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATG GTGGCGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GTGGCGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGGGACC TCGGGCGACACG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACACG							•
5151 AGGGACGCTG AGCACCTTGG GTGGTGTGAT CCTGAGCCCC GGCTTCCCAG 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTTTT TGAAAACAGA GGTAAGGACG 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATTGGAGA AGAGTAATTC ACGCAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATTGGAGAA AGAGTAATTC ACGCAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATTGGAGAA AGAGTAATTC ACGCAGAGCCG 5701 ACTTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGCCCG GCCGCGGGGG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GCCGCGGTGG 5801 CATCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCCGCGGTGG 5801 CATCACGCCTG TAATCCCAGC ACTTTGGAAG CCCGAGGCGG GCCGCGGTGG 5801 CATCACGCCTG TAATCCCAGC ACTTTGGAAG CCCGAGGCGG GCCGCGGTGG 5801 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGGCG CCTGTAGTCC 5901 AGGTCAGGG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6001 AGCTTCCGG GAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6001 AGCTTCCGG GAGGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6001 AGCTTCCAGGT GAGGAGAGAT CGCCCACTG CACCTCCAGCC TGGCCGACGCG	•						-
45 5201 GTTCTTACCC CAACAACTTA GACTGCACCT GGAGGATCTC ATTACCCATC 5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACACA ACGAGAACCG AAACCCAAAT 5451 CTAAATACAC TTCTTACATG TAAATTGTAT TTAAGTATAA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTTTT TGAAAACAGA GGTAAGCAC 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGCCG GGCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGGTGG 5851 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG CCTGTAGTCC 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG CCTGTAGTCC 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG CCTGTAGTCC 60 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGCGG 6051 AGCTTGCAGT GAGGAGAAT CGCCGCCACTG CACTCCAGCC TGGGCGACAG							
5251 GGCTATGGTG CACATATTCA GTTTCTGAAT TTTTCTACCG AAGCTAATCA 5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACACA ACGAGAACCG AAACCCAAAT 5551 CTGGTTCCAA GCTTGTACAG GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ACTTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GCGCGGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGC CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACACG							
5301 TGACTTCCTT GAAATTCAAA ATGGACCTTA CCACACCAGC CCCATGATTG 5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5501 CTAAATACAC TTCTTACATG TAAATTGTAT TTAAGTATAA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCCG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACACG	45						
5351 GACAATTTAG CGGCACGGAT CTCCCCGCGG CCCTGCTGAG CACAACGCAT 5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5501 CTAAATACAC TTCTTACATG TAAATTGTAT TTAAGTATAA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCCG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACACG	45						
5401 GAAACCCTCA TCCACTTTTA TAGTGACCAT TCGCAAAACC GGCAAGGATT 5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5501 CTAAATACAC TTCTTACATG TAAATTGTAT TTAAGTATAA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCCG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG		5301					
5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5501 CTAAATACAC TTCTTACATG TAAATTGTAT TTAAGTATAA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG		5351	GACAATTTAG	CGGCACGGAT	CTCCCCGCGG	CCCTGCTGAG	CACAACGCAT
5451 TAAACTTGCT TACCAAGNTA TGGAACAACA ACGAGAACCG AAACCCAAAT 5501 CTAAATACAC TTCTTACATG TAAATTGTAT TTAAGTATAA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG		5401	GAAACCCTCA	TCCACTTTTA	TAGTGACCAT	TCGCAAAACC	GGCAAGGATT
5501 CTAAATACAC TTCTTACATG TAAATTGTAT TTAAGTATAA ATCTCCCTAA 5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG							
5551 CTGGTTCCAA GCTTGTACGA GTGGAATAAT TTTTTGGTGG AATGTTGGTT 5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG	50						
5601 TCTGGTTAGT AGTGGAACAC TTGTTGTTTT TGAAAACAGA GGTAAGGACA 5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5701 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG							
5651 CAGACGGAAC CACCAGTGGG TTCGCCTTTT CTGCTGCCCA GACAGAGCCG 5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5751 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGCCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG							
5701 ATTTATCAAG ACGGGAATTG CAATGGAGAA AGAGTAATTC ACGCAGAGCC 5751 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG							
55 5751 AGATGTGTGG GAGACCGGAG TTTTATTGTG ACTCAATTCA GTCTCCCCAG 5801 CATTCAGGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG							
5801 CATTCAGGA TTCAAGTTTT TAAAGATAAT TTGGCGGCCG GGCGCGGTGG 5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG							
5851 CTCACGCCTG TAATCCCAGC ACTTTGGAAG GCCGAGGCGG GCGGATCACG 5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG)						
5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG		5801	CATTCAGGGA	TTCAAGTTTT	TAAAGATAAT	TTGGCGGCCG	GGCGCGGTGG
5901 AGGTCAGGAG ATCGAGACCA TCCTGGCTAA CACGGTGAAA CCCCGTCTCT 5951 ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG			CTCACGCCTG	TAATCCCAGC	ACTTTGGAAG	GCCGAGGCGG	GCGGATCACG
5951 ACTAAAAATA CCAAAAATTA GCCGGGCATA GTGGCGGGCG CCTGTAGTCC 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG							
60 6001 CAGCTACTCG GGAGGCTGAG GCAGGANAGT GGCGTGAACC CGGGAGGCGG 6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG	•						
6051 AGCTTGCAGT GAGGAGAGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG	60						
biui agccagacte catetegaaa aaaaaaaaaa aaaaaaaaa aaaaa	•						
:	•	P101	AGCCAGACTC	CATCTCGAAA	Ададада	AAAAAAAAA	AAAAA

SEQ ID NO:3

Brank Service	EC 2372 No.	cleotide se	mionao	6409 bp		
	1		GGTATGAATT		ጥ ልርጥልጥጥል <i>ር</i> አ	מיד ביד ביד ביד ביד ביד ביד ביד ביד ביד ב
5	51		TCCAAGTGAA			
٠	101		CCCAGCAAGG			
	151		TCAACCATGT			
	201		CCATGGCTTC			
1.4	' 251		CTCTGACATG			
10	301		GTTCCGACTT			
10.	351		TACGTGCTCC			
	401		GCTCGCTGCT			
	451		GATCCAATCT			
	501		GTTCAGTATG			
15	551		CCCGGACAAG			
12	601		GCTATGACAC			
	651		TCGGTCTTGT			
	701		CATGAGCAAC			
	751		CACCTGGGTT			
20	801		GATCCTGGAA			
	· 851		TGGAGATACA			
	901		AGAGAGTTAT			
	951		AGCTGTGTAT			
	1001		TCTGTCACCA			
25	1051		GGTTGATTAT			
	1101		TTTGATGTTG			
	1151		TTCTGACATA			
	1201		AGCTGGCCAG			
	1251		TCCACTACTG			
30	1301	TTGGTCAGAA	TGAGTGCCAT	GATCCTGGCA	TTCCTATAAA	CGGACGACGT
	1351		GGTTTCTACT			
	1401		AAGACCCAGG			
	1451	ACGGGAACGT	GGTCTGGAGC	TCCACCGTGC	CCCGCTGTGA	AGCTCCATGT
	1501		TGACAGCGTC			
35	1551		AAGGATTCTT			
	1601	CAGGCCACTC	TATCAAAATA	ACTTTTGACA	GATTTCAGAC	AGAGGTCAAT
	1651	TATGACACCT	TGGAGGTCAG	AGATGGGCCA	GCCAGTTCGT	CCCCACTGAT
	1701	CGGCGAGTAC	CACGGCACCC	AGGCACCCCA	GTTCCTCATC	AGCACCGGGA
40	1751		CCTGCTATTC			
40	1801	TTCCTCATCC	ACTATGAGAG	TGTGACGCTT	GAGTCGGATT	CCTGCCTGGA
	1851		CCTGTGAACG			
	1901		GACTTTCAGC			
	1951		TCTGTGAGAG			
45	2001		CTATGTGGAG GTTTCCAGAT			
43	2051	TITUTUUTGG	TGTCTCATGG	CANACCACTO	CANAGEMENT	TTCACACGIGG
	2101	MCATTGAAG	AGTTCCCACG	ACMAMOGAGII	CAMMIGATOL	CATCCACCCII
	2151 2201	TCATCTTGAG	CGTTGCCACG	CTCACCCCCT	CCCTCTTCCC	TCDTDCCDTC
	2251	NACCCACCCC	TGTTNGGAAA	CTCACCGGGT	CACCUTCGGT	ΤΕΛΙΛΟΘΑΙΟ
50	2301	CONCOCARO	TCGTACGAGG	CITCACIGCC	CAGCIICGGI	CANTATCACA
50	2351	TECACCOATC	TGATGATCCT	CCACTCCCTC	CCTTCACCC	AAGAATTGGT
	2401	TUGHUCATU	GTGTGGGAGA	CTCTCTCACA	TTTTCAGCCG	TCCTGGGGTTA
	2451		GGTGCCACCA			
	2501	TCTCCACTCC	ACCTCTGCCA	AGCTCTGTGG	CCGAATGTGG	AGCAAGTGTC
55	2551	AAACCAAATC	AAGGAACATT	ACTETETECA	AATTTTCCAT	CCAATTATGA
55	2601	TANGGRAMIG	GAGTGTATCT	ATAAAATAGA	AACAGAAGCC	GGCAAGGGCA
	2651	TOPIANCONI	AACACGAAGC	TTCCAGCTGT	TTGAAGGAGA	TACTCTAAAG
	2701		GAAAAGACAG			
	2751		CTGGGGCTGA			
60	2801	TAGAGTTCAA	CACCAATGGA	TCTGACACCG	ACCAAGGTTT	TCAACTCACC
	2851	TATACCAGTT	TTGATCTGGT	AAAATGTGAG	GATCCGGGCA	TCCCTAACTA
	2901	CGGCTATAGG	ATCCGTGATG	AAGGCCACTT	TACCGACACT	GTAGTTCTGT
	2951	ACAGTTGCAA	CCCGGGGTAC	GCCATGCATG	GCAGCAACAC	CCTGACCTGT

				COCCCO CD D D	OCT ONT COMM	CCMCCAMACC
	3001	TTGAGTGGAG	ACAGGAGAGT	GTGGGACAAA	CCACTACCTT	CGIGCAIAGC
	3051	GGAATGTGGT	GGTCAGATCC	ATGCAGCCAC	ATCAGGACGA	ATATTGTCCC
	3101	CTGGCTATCC	AGCTCCGTAT	GACAACAACC	TCCACTGCAC	CTGGATTATA
_	3151	GAGGCAGACC	CAGGAAAGAC	CATTAGCCTC	CATTTCATTG	TTTTCGACAC
5	3201				GGACGGGCCG	
	3251	ACATCCTGCT	GAAGGAGTGG	AGTGGCTCCG	CCCTTCCGGA	GGACATCCAC
	3301	AGCACCTTCA	ACTCACTCAC	CCTGCAGTTC	GACAGCGACT	TCTTCATCAG
	3351	CAAGTCTGGC	TTCTCCATCC	AGTTCTCCAC	CTCAATTGCA	GCCACCTGTA
•	3401	ACGATCCAGG	TATGCCCCAA	AATGGCACCC	GCTATGGAGA	CAGCAGAGAG
1.0	3451	GCTGGAGACA	CCGTCACATT	CCAGTGTGAC	CCTGGCTATC	AGCTCCAAGG
	3501	ACAAGCCAAA	ATCACCTGTG	TGCAGCTGAA	TAACCGGTTC	TTTTGGCAAC
	3551	CAGACCCTCC	TACATGCATA	GCTGCTTGTG	GAGGGAATCT	GACGGGCCCA
	3601	GCAGGTGTTA	TTTTGTCACC	CAACTACCCA	CAGCCGTATC	CTCCTGGGAA
	3651	GGAATGTGAC	TGGAGAGTAA	AAGTGAACCC	GGACTTTGTC	ATCGCCTTGA
15	3701	TATTCAAAAG	TTTCAACATG	GAGCCCAGCT	ATGACTTCCT	ACACATCTAT
	3751	GAAGGGGAAG	ATTCCAACAG	CCCCCTCATT	GGGAGTTACC	AGGGCTCTCA
	3801	GGCCCCAGAA	ACAATACACA	GTAGCGGAAA	CAGCCTGTTT	CTGGCATTTC
	3851	CCACTCATCC	CTCCCTCCCC	CTTTCAGGGT	TCGCCATTGA	ATTTAAAGAG
	3901	ANACCACEC	AACCTTCTTT	TCACCCAGGA	AATATAATGA	ATGGGACAAG
20		ACCEPT ACA	CACOUCTATE	TOMOCOMOCIA	CATCACCTAC	CAGTGTGACT
20	3951	AGII GGAACA	CARROCTECACOC	CCCTCDTCCA	TCACCTGTGT	CATTCCCCCT
	4001	CIGGCIAIAA	COMCOMCCO	CCTCATCCA	CCCTCCTGCA	ATTCCTCCCTC
•	4051	GATGGGAAAC	CCTCCTGGGA	CAAGIGCIG	A COUNTRY OF A	CCAAACTACCIG
	4101	TGGAGGCCAG	TACACGGGAT	CAGAAGGGI	AGTTTTATCA	CDAAAACIACC
0.5	4151	CCCATAATTA	CACAGCTGGT	CAAATATGCC	TCTATTCCAT	CACGGIACCA
25	4201	AAGGAATTCG	TGGTCTTTGG	ACAGTTTGCC	TATTTCCAGA	CAGCCCTGAA
	4251	TGATTTGGCA	GAATTATTTG	ATGGAACCCA	TGCACAGGCC	AGACTTCTCA
	4301	GCTCACTCTC	GGGGTCTCAC	TCAGGGGAAA	CATTGCCCTT	GGCTACGTCA
	4351	AATCAAATTC _.	TGCTCCGATT	CAGTGCAAAG	AGCGGTGCCT	CTGCCCGCGG
	4401	CTTCCACTTC	GTGTATCAAG	CTGTTCCTCG	TACCAGTGAC	ACCCAATGCA
30	4451	GCTCTGTCCC	CGAGCCCAGA	TACGGAAGGA	GAATTGGTTC	TGAGTTTTCT
	4501	GCCGGCTCCA	TCGTCCGATT	CGAGTGCAAC	CCGGGATACC	TGCTTCAGGG
	4551	TTCCACGGCG	CTCCACTGCC	AGTCCGTGCC	CAACGCCTTG	GCACAGTGGA
	4601	ACGACACGAT	CCCCAGCTGT	GTGGTACCCT	GCAGTGGCAA	TTTCACTCAA
	4651	CGAAGAGGTA	CAATCCTGTC	CCCCGGCTAC	CCTGAGCCAT	ACGGAAACAA
35	4701	CTTGAACTGT	ATATGGAAGA	TCATAGTTAC	GGAGGGCTCG	GGAATTCAGA
	4751	TCCAAGTGAT	CAGTTTTGCC	ACGGAGCAGA	ACTGGGACTC	CCTTGAGATC
	4801	CACGATGGTG	GGGATGTGAC	CGCACCCAGA	CTGGGAAGCT	TCTCAGGCAC
	4851	CACAGTACCG	GCACTGCTGA	ACAGTACTTC	CAACCAACTC	TACCTGCATT
	4901	TCCAGTCTGA	CATTAGTGTG	GCAGCTGCTG	GTTTCCACCT	GGAATACAAA
40	4951	ACTGTAGGTC	TTGCTGCATG	CCAAGAACCA	GCCCTCCCCA	GCAACAGCAT
	5001	CAAAATCGGA	GATCGGTACA	TGGTGAACGA	CGTGCTCTCC	TTCCAGTGCG
	5051	AGCCCGGGTA	CACCCTGCAG	GGCCGTTCCC	ACATTTCCTG	TATGCCAGGG
	5101	ACCGTTCGCC	GTTGGAACTA	TCCGTCTCCC	CTGTGCATTG	CAACCTGTGG
	5151	AGGGACGCTG	ACCACCTTCC	GTGGTGTGAT	CCTGAGCCCC	GGCTTCCCAG
45	5201	CTTCTTACCC	CDACAACTTA	GACTGCACCT	GGAGGATCTC	ATTACCCATC
43	5251	CCCTATCCTC	CACATATTCA	CTTTCTCAAT	TTTTCTACCG	AAGCTAATCA
	. 5301	TCACTTCCTT	CADATTCADA	DTCCDCCTTA	CCACACCAGC	CCCATGATTG
		CACAAMMMAAC	CCCCACCCAT	CTCCCCCCC	CCCTGCTGAG	CACAACGCAT
	5351	CANACCCUCA	TOGCA COUNTY	TACTCACCAT	TCGCAAAACC	CCCDACCATT
50	5401	GAAACCCICA maaacconcco	UNCCANCOCA	TAGIGACCAI	GAACTGTCCA	CATCCACCCC
50	5451	TAAACTTGCT	TACCAAGCCT	ATGAATTACA	AMMACACCCM	CCCCCAATCA
	5501	CATTTCAGAA	TGGGTACATG	ATCAACTCGG	ATTACAGCGT	A WCCMCMCCCM
	5551	GTATCTTTCG	AGTGTTATCC	TGGGTACATT	CTAATAGGCC	ATCCIGICCI
	. 5601	CACTTGTCAG	CATGGGATCA	ACAGAAACTG	GAACTACCCT	TTTCCAAGAT
	5651	GTGATGCCCC	TTGTGGGTAC	AACGTAACTT	CTCAGAACGG	CACCATCTAC
55	5701	TCCCCTGGCT	TTCCTGATGA	GTATCCGATC	CTGAAGGACT	GCATTTGGCT
	5751	CATCACGGTG	CCTCCAGGGC	ACGGAGTTTA	CATCAACTTC	ACCCTGTTAC
	5801	AGACGGAAGC	TGTCAACGAT	TACATTGCTG	TTTGGGACGG	TCCCGATCAG
	5851	AACTCACCCC	AGCTGGGAGT	TTTCAGTGGC	AACACAGCCC	TCGAAACGGC
	5901	GTATAGCTCC	ACCAACCAAG	TCCTGCTCAA	GTTCCACAGC	GACTTTTCAA
60	5951	ATGGAGGCTT	CTTTGTCCTC	AATTTCCACG	GTCAGTTGAT	TTTCACTCCG
	6001	TTAGTTAAGA	CTGAGAATTC	CATGTGGTGT	TTACTGCAGT	GTTGTCCCAC
	6051	GCCTTGTTTC	CAGCTGAAGT	TTCTTGATTC	AGCCGAGGGC	GTGTATGATT
	6101	CTTTTGCACT	GGAGGCCAGC	GTTTCCTGTG	GTCCTTTTTT	TGTTTAATGA
	0101	JIIIIOMOI	30			

620 625 630 5 635		GTGAAACTCT ATTTACTCAT CTATAAATGG	AAGATGAAGA CCCTGTCTCA TGTGAAAGCA	CCATTGAAAG AGATAAGGTG AACCTCCAAT	AGATTTGGTA TTATAGCAAA AATCCTGGGA
----------------------------	--	--	--	--	--

SEQ ID NO:4

: un	EC. 2372 No.	Santida san		5667 bp		
	3G-3V3 Nu	cleotide sec			TAGTATTACA	Δ ጥ Δ ጥ Δ ጥ Δ ጥ Δ
5	51				GAGTTGAAGT	
	101	CANCARCO	CCCAGGGAA	ATCCARCCA	TAAAAACTCT	CTCTCCCATA
	151				GAAGATACGT	
	201				TCAAGTGAGC	
* 1	251				CTGGGATTCC	
10	301	DCDDCDCCDC	CTTCCCACTT	CAGGGTTGGT	GCAAATGTAC	AGTTTTCATG
1,0,	351				AAGCATCACC	
	401	TTACAGAGAC	GCTCGCTGCT	TGGAGTGACC	ACAGGCCCAT	CTGCCGAGCG
	451	AGAACATGTG	GATCCAATCT	GCGTGGGCCC	AGCGGCGTCA	TTACCTCCCC
	501				ACACTGTGTG	
15	551	CCACCACCGA	CCCGGACAAG	GTCATCAAGC	TTGCCTTNGA	AGAGTTTGAG
	601	CTGGAGCGAG	GCTATGACAC	CCTNACGGTT	GGTGATGCTG	GGAAGGTGGG
	651	AGACACCAGA	TCGGTCTTGT	ANGTGCTCAC	GGGATCCAGT	GTTCCTGACC
	701	TCATTGTGAG	CATGAGCAAC	CAGATGTGGC	TACATCTGCA	GTCGGATGAT
	751	AGCATTGGCT	CACCTGGGTT	TAAAGCTGTT	TACCAAGAAA	TTGAAAAGGG
20	801	AGGGTGTGGG	GATCCTGGAA	TCCCCGCCTA	TGGGAAGCGG	ACGGGCAGCA
	851	GTTTCCTCCA	TGGAGATACA	CTCACCTTTG	AATGCCCGGC	GGCCTTTGAG
	901				CAGAACAATC	
	951				CTTCAACTTT	
	1001	CTGGGATTAT	TCTGTCACCA	AATTATCCAG	AGGAATATGG	GAACAACATG
25	1051				GGAAGTCGAA	
	1101	CTTTAATGAT	TTTGATGTTG	AGCCTCAATT	TGACTTTCTC	GCGGTCAAGG
	1151				GTACTTTTTC	
	1201	GTGCCTTCCC	AGCTGGCCAG	CAGTGGGCAT	ATAGTTCGCT	TGGAATTTCA
	1251	GTCTGACCAT	TCCACTACTG	GCAGAGGGTT	CAACATCACT	TACACCACAT
30	1301	TTGGTCAGAA	TGAGTGCCAT	GATCCTGGCA	TTCCTATAAA	CGGACGACGT
	1351	TTTGGTGACA	GGTTTCTACT	CGGGAGCTCG	GTTTCTTTCC	ACTGTGATGA
	1401	TGGCTTTGTC	AAGACCCAGG	GATCCGAGTC	CATTACCTGC	ATACTGCAAG
	1451	ACGGGAACGT	GGTCTGGAGC	TCCACCGTGC	CCCGCTGTGA	AGCTCCATGT
	1501	GGTGGACATC	TGACAGCGTC	CAGCGGAGTC	ATTTTGCCTC	CTGGATGGCC
35	1551	AGGATATTAT	AAGGATTCTT	TACATTGTGA	ATGGATAATT	GAAGCAAAAC
	1601	CAGGCCACTC	TATCAAAATA	ACTTTTGACA	GATTTCAGAC	AGAGGTCAAT
	1651	TATGACACCT	TGGAGGTCAG	AGATGGGCCA	GCCAGTTCGT	CCCCACTGAT
	1701	CGGCGAGTAC	CACGGCACCC	AGGCACCCCA	GTTCCTCATC	AGCACCGGGA
40	1751	AÇTTCATGTA	CCTGCTATTC	ACCACTGACA	ACAGCCGCTC	CAGCATCGGC
40	1801	TTCCTCATCC.	ACTATGAGAG	TGTGACGCTT	GAGTCGGATT	DDDCCCDDCA
	1851	CCCGGGCATC	CCTGTGAACG	GCCATCGCCA	CGGTGGAGAC GGTACACACT	A A CTC A CCAC
•	1901	GGTCCACAGT	GACTTTCAGC	TGTGACCCGG	TGGAACCACG	CCTTCCCCAC
	1951	GAGCCCCTCG	TUTGTGAGAG	CCTACATCCA	AGGGAAGAGT	CCTIGCCCAG
45	2001 2051	MMMCMCCMCC	CTATGIGGAG	TOTAL COA	ACTCTCTAAA	CTCCACCTCC
45	2101	ACCAMMCAAC	TOTO TO THE	CANACCACTT	CAAATGATCT	TTCACACCTT
	2151	TCATCTTCAG	ACTTCCCACG	DCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	GATCACAGAG	GATGGAAGTT
	2201	TURICIIGAG.	CCTTCCACG	CTCACCGGGT	CGGTGTTGCC	TCATACGATC
	2251	AAGGCAGGCC	TETTNEGADA	CTTCACTGCC	CAGCTTCGGT	TTATATCAGA
50	2301		TCGTACGAGG	CCTTCAATAT	CACATTTTCA	GAATATGACC
50	2351	TECACCCATE	TGATGATCCT	GGAGTCCCTG	CCTTCAGCCG	AAGAATTGGT
	2401	TUCKCCATC	GTGTGGGAGA	CTCTCTGACG	TTTTCCTGCT	TCCTGGGATA
	2451	TCCTTTAGAA	GETGCCACCA	AGCTTACCTG	CCTGGGTGGG	GGCCGCCGTG
	2501	TETEGAGTEC	ACCTCTGCCA	AGGTGTGTGG	CCGAATGTGG	AGCAAGTGTC
55	2551	AAAGGAAATG	AAGGAACATT	ACTGTCTCCA	AATTTTCCAT	CCAATTATGA
22	2601	TAATAACCAT	GAGTGTATCT	ATAAAATAGA	AACAGAAGCC	GGCAAGGGCA
	2651	TCCACCTTAG	AACACGAAGC	TTCCAGCTGT	TTGAAGGAGA	TACTCTAAAG
	2701	GTATATGATG	GAAAAGACAG	TTCCTCACGT	CCACTGGGCA	CGTTCACTAA
	2751	AAATGAACTT	CTGGGGCTGA	TCCTAAACAG	CACATCCAAT	CACCTGTGGC
60	2801	TAGAGTTCAA	CACCAATGGA	TCTGACACCG	ACCAAGGTTT	TCAACTCACC
	2851	TATACCAGTT	TTGATCTGGT	AAAATGTGAG	GATCCGGGCA	TCCCTAACTA
	2901	CGGCTATAGG	ATCCGTGATG	AAGGCCACTT	TACCGACACT	GTAGTTCTGT
	2951	ACAGTTGCAA	CCCGGGGTAC	GCCATGCATG	GCAGCAACAC	CCTGACCTGT

				7 G 7 G C 7 C 7 C 70	CHCCCACAAA	ここれと中本とと中中	CCTCCATACC
		3001	TTGAGTGGAG	ACAGGAGAG1	GTGGGACAAA ATGCAGCCAC	DUCYCCYCCY COA	ATATTCTCCC
		3051	GGAATGTGGT	GGTCAGATCC	GACAACAACC	MICAGGACGA MCCACTCCAC	CTCCATTATA
: 535:	• ;	3101	CTGGCTATCC	AGCTCCGTAT	CATTAGCCTC	CATTOCACTOCAC	TTTTTTTTT
_		3151	GAGGCAGACC	CAGGAAAGAC	TATTAGCCTC	CCACCCCCCC	CTCCACACAC
5		3201	GGAGATGGCT	CACGACATCC	TCAAGGTCTG	CCCTTCCCCA	CCACATCCAC
•		3251	ACATCCTGCT	GAAGGAGTGG	AGTGGCTCCG CCTGCAGTTC	CACACCCACT	TCTTCATCAC
		3301	AGCACCTTCA	ACTCACTCAC	CCTGCAGTTC	CMCA A TTCCA	CCCACCACCA
		3351	CAAGTCTGGC	TTCTCCATCC	AGTTCTCCAC	CICAMIIGCA	CACCACACAC
* ***		3401	ACGATCCAGG	TATGCCCCAA	AATGGCACCC	CCTAIGGAGA	ACCTCCA ACC
10		3451	GCTGGAGACA	CCGTCACATT	CCAGTGTGAC	CCIGGCIAIC	MUCICCAAGG
		3501	ACAAGCCAAA	ATCACCTGTG	TGCAGCTGAA	CACCCAARCE	LILIGGCAAC
		3551	CAGACCCTCC	TACATGCATA	GCTGCTTGTG	CACCCCTATC	CECCECCON
		3601	GCAGGTGTTA	TTTTGTCACC	CAACTACCCA	CAGCUGIAIC	AUCCCCUTTCA
		3651	GGAATGTGAC	TGGAGAGTAA	AAGTGAACCC	AMCA COMCCOM	AICGCCIIGA
15		3701	TATTCAAAAG	TTTCAACATG	GAGCCCAGCT	ATGACTICCI	ACACATCTAT
		3751	GAAGGGGAAG	ATTCCAACAG	CCCCCTCATT	GGGAGTTACC	AGGGCTCTCA
		3801	GGCCCCAGAA	AGAATAGAGA	GTAGCGGAAA	CAGCCTGTTT	DEGLATTIC
		3851	GGAGTGATGC	CTCCGTGGGC	CTTTCAGGGT	TCGCCATTGA	ATTTAAAGAG
		3901	AAACCACGGG	AAGCTTGTTT	TGACCCAGGA	AATATAATGA	ATGGGACAAG
20		3951	AGTTGGAACA	GACTTCAAGC	TTGGCTCCAC	CATCACCTAC	CAGTGTGACT
		4001	CTGGCTATAA	GATTCTTGAC	CCCTCATCCA	TCACCTGTGT	CATTGGGGCT
		4051	GATGGGAAAC	CCTCCTGGGA	CCAAGTGCTG	CCCTCCTGCA	ATGCTCCCTG
		4101	TGGAGGCCAG	TACACGGGAT	CAGAAGGGGT	AGTTTTATCA	CCAAACTACC
		4151	CCCATAATTA	CACAGCTGGT	CAAATATGCC	TCTATTCCAT	CACGGTACCA
25		4201	AAGGAATTCG	TGGTCTTTGG	ACAGTTTGCC	TATTTCCAGA	CAGCCCTGAA
		4251	TGATTTGGCA	GAATTATTTG	ATGGAACCCA	TGCACAGGCC	AGACTTCTCA
		4301	GCTCACTCTC	GGGGTCTCAC	TCAGGGGAAA	CATTGCCCTT	GGCTACGTCA
		4351	AATCAAATTC	TGCTCCGATT	CAGTGCAAAG	AGCGGTGCCT	CTGCCCGCGG
		4401	CTTCCACTTC	GTGTATCAAG	CTGTTCCTCG	TACCAGTGAC	ACCCAATGCA
30		4451	GCTCTGTCCC	CGAGCCCAGA	TACGGAAGGA	GAATTGGTTC	TGAGTTTTCT
		4501	GCCGGCTCCA	TCGTCCGATT	CGAGTGCAAC	CCGGGATACC	TGCTTCAGGG
		4551	. TTCCACGGCG	CTCCACTGCC	AGTCCGTGCC	CAACGCCTTG	GCACAGTGGA
		4601	ACGACACGAT	CCCCAGCTGT	GTGGTACCCT	GCAGTGGCAA	TTTCACTCAA
		4651	CGAAGAGGTA	CAATCCTGTC	CCCCGGCTAC	CCTGAGCCAT	ACGGAAACAA
35		4701	CTTGAACTGT	ATATGGAAGA	TCATAGTTAC	GGAGGGCTCG	GGAATTCAGA
		4751	TCCAAGTGAT	CAGTTTTGCC	ACGGAGCAGA	ACTGGGACTC	CCTTGAGATC
		4801	CACGATGGTG	GGGATGTGAC	CGCACCCAGA	CTGGGAAGCT	TCTCAGGCAC
		4851	CACAGTACCG	GCACTGCTGA	ACAGTACTTC	CAACCAACTC	TACCTGCATT
		4901	TCCAGTCTGA	CATTAGTGTG	GCAGCTGCTG	GTTTCCACCT	GGAATACAAA
40		4951	ACTGTAGGTC	TTGCTGCATG	CCAAGAACCA	GCCCTCCCCA	GCAACAGCAT
		5001	CAAAATCGGA	GATCGGTACA	TGGTGAACGA	CGTGCTCTCC	TTCCAGTGCG
		5051	AGCCCGGGTA	CACCCTGCAG	GGCCGTTCCC	ACATTTCCTG	TATGCCAGGG
		5101	ACCGTTCGCC	GTTGGAACTA	TCCGTCTCCC	CTGTGCATTG	CAACCTGTGG
		5151	AGGGACGCTG	AGCACCTTGG	GTGGTGTGAT	CCTGAGCCCC	GGCTTCCCAG
45		5201	GTTCTTACCC	CAACAACTTA	GACTGCACCT	GGAGGATCTC	ATTACCCATC
		5251	CCCTATCCTC	CACATATTCA	GTTTCTGAAT	TTTTCTACCG	AAGCTAATCA
		5301	TGACTTCCTT	GAAATTCAAA	ATGGACCTTA	CCACACCAGC	CCCATGATTG
		5351	CACAATTTAG	CGGCACGGAT	CTCCCCGCGG	CCCTGCTGAG	CACAACGCAT
•		5401	CAAACCCTCA	TCCACTTTTA	TAGTGACCAT	TCGCAAAACC	GGCAAGGATT
50		5451	ም ል ል ል ር ጥጥር ር ጥ	TACCAAGCCT	AATCTGGAAA	CATTGGTCCT	GCTTTCCCAT
		5501	GTCTTGACAC	CCCATTCCAA	GCCAGATGTC	AAGGAGAAGA	AAGGACTTTC
		5551	ΑΔΑΔΑΔΥΤΔΔ	AAAACAAAAA	CTCGAAACAA	CATGTTTTT	ATTGTACGCC
		5601	ATTAATTTCC	TATCACTGAG	ATATAAAAAT	AAATAATGCC	AAAAAAAA
		5651	AAAAAAAAA				
55							

SEQ ID NO:5

	5R-3V2 Nu	cleotide sed	quence	7323 bp .		
	1		CGCGGCGGGT		GGCNCTCTCT	CCGGCTCGCC
5	51		GTGATTATTT			
-	101		GTGTCGCGTG			
	151		. CCAGTCGCTG			
	201		TCACTGCAGC			
	251	GGGTCCCAAT	GGCACTATTG	AGAGCCCAGG	GTTTCCTCAC	GGGTATCCGA
10	301		CTGCACCTGG			
	351		TCCATACCTT			
	401		GGACAGCCTC			
	451		GCCCTCCTCT			
	501	TGGTTCACGA	CAGACTTCGC	TGTGAGTGCC	CAAGGTTTCA	AAGCATTATA
15	551		CCTAGCCACA			
	601		TGGAACGAGA			
	651		GCTACATCTT			
7	701		AATGGTGCAT			
	751		CTGCGGAGGA			
20	801		TCCCTTCAGA			
	851	CATTCTGGCT	GAGCCCGGGG	ACACCATTGC	GCTGGTCTTC	ACTGACTTTC
	901		AGGATATGAT			
	951		TAACTGGCAT			
	1001	GAATTGGCTA	CGACTCCATT	TCACCTCTGA	CAGCAACCAC	CGACGCAAAG
25	1051	GATTTAACGC	TCAGTTCCAA	GTGAAAAAGG	CGATTGAGTT	GAAGTCAAGA
	1101	GGAGTCAAGA	TGCTGCCCAG	CAAGGATGGA	AGCCATAAAA	ACTCTGTCTT
	1151	GAGCCAAGGA	GGTGTTGCAT	TGGTCTCTGA	CATGTGTCCA	GATCCTGGGA
	1201	TTCCAGAAAA	TGGTAGAAGA	GCAGGTTCCG	ACTTCAGGGT	TGGTGCAAAT
	1251	GTACAGTTTT	CATGTGAGGA	CAATTACGTG	CTCCAGGGAT	CTAAAAGCAT
30	1301	CACCTGTCAG	AGAGTTACAG	AGACGCTCGC	TGCTTGGAGT	GACCACAGGC
	1351		AGCGAGAACA			
	1401	GTCATTACCT	CCCCTAATTA	TCCGGTTCAG	TATGAAGATA	ATGCACACTG
	1451	TGTGTGGGTC	ATCACCACCA	CCGACCCGGA	CAAGGTCATC	AAGCTTGCCT
	1501	TNGAAGAGTT	TGAGCTGGAG	CGAGGCTATG	ACACCCTNAC	GGTTGGTGAT
35	1551	GCTGGGAAGG	TGGGAGACAC	CAGATCGGTC	TTGTANGTGC	TCACGGGATC
	. 1601		GACCTCATTG			
	1651		TGATAGCATT			
	1701		AGGGAGGGTG			
40	1751		AGCAGTTTCC			
40	1801		TGAGCTGGTG			
	1851		CTGGCAACAA			
	1901		TCATCTGGGA			
	1951		CATGAACTGT			
45	2001		TAATCTTTAA			
45	2051		AAGGATGATG			
	2101		TGAAGTGCCT			
	2151		TTCAGTCTGA			
	2201		ACNTTTGGTC			
50	2251		ACGTTTTGGT			
50	2301		ATGATGGCTT CAAGACGGGA			
	2351		ATGTGGTGGA			
	2401		GGCCAGGATA			
	2451		AAACCAGGCC			
55	2501		CAATTATGAC			
55	2551 2601		TGATCGGCGA			
	2601 2651		GGGAACTTCA			
	2651 . 2701		CGGCTTCCTC			
			TGGACCCGGG			
60	2751 2801		ATCAGGTCCA			
00	2851		CGACGAGCCC			
	2901		CCAGCTGCGA			
	2951		GTCCTTTCTC			
	7 2 3 T	OVO I GOWACH	GICCILICIC	0100011100.		00122101010

				CT T COCOCOCO	አ ሞርርር እ አ አ ርር	አርጥጥር <u>አአአጥር</u>
	3001	TAAACTGCAC	GTGGACCATT	GAAGIGICIC	CACCACRARGO	MOIICAAAIG
	3051	ATCTTTCACA	CCTTTCATCT	TGAGAGTTCC	CACCACIAII	CCCTCCCTCT
	3101	AGAGGATGGA	AGTTTTTCCG	AGCCCGTTGC	CAGGCTCACC	TCCCCACCAA
_	3151	TGCCTCATAC	GATCAAGGCA	GGCCTGTTNG	CACCCCTTCAC	TGCCCAGCTT
5	3201	CGGTTTATAT	CAGACTTCTC	AATTTCGTAC	TO THE TOTAL OF TH	CCTCCCTTCA
	3251	TTCAGAATAT	GACCTGGAGC	CATGTGATGA.	TCCTGGAGTC	COLGCUITCA
	3301	GCCGAAGAAT	TGGTTTTCAC	TTTGGTGTGG	GAGACTETET	GACGTTTTCC
	3351	TGCTTCCTGG	GATATCGTTT	AGAAGGTGCC	ACCAAGCTTA	CCTGCCTGGG
	3401	TGGGGGCCGC	CGTGTGTGGA	GTGCACCTCT	GCCAAGGTGT	GTGGCCGAAT
10	3451	GTGGAGCAAG	TGTCAAAGGA	AATGAAGGAA	CATTACTGTC	TCCAAATTTT
	3501	CCATCCAATT	ATGATAATAA	CCATGAGTGT	ATCTATAAAA	CHCHARCAGA
	3551	AGCCGGCAAG	GGCATCCACC	TTAGAACACG	AAGCTTCCAG	A COMOCA COC
	3601	GAGATACTCT	AAAGGTATAT	GATGGAAAAG	ACAGTTCCTC	ACGICCACIG
_	3651	GGCACGTTCA	CTAAAAATGA	ACTTCTGGGG	CTGATCCTAA	ACAGCACAIC
15	3701	CAATCACCTG	TGGCTAGAGT	TCAACACCAA	TGGATCTGAC	ACCGACCAAG ·
	3751	GTTTTCAACT	CACCTATACC	AGTTTTGATC	TGGTAAAATG	TGAGGATCCG
	3801	GGCATCCCTA	ACTACGGCTA	TAGGATCCGT	GATGAAGGCC	ACTITACCGA
· .	3851	CACTGTAGTT	CTGTACAGTT	GCAACCCGGG	GTACGCCATG	CATGGCAGCA
•	3901	ACACCCTGAC	CTGTTTGAGT	GGAGACAGGA	GAGTGTGGGA	CAAACCACTA
20	3951	CCTTCGTGCA	TAGCGGAATG	TGGTGGTCAG	ATCCATGCAG	CCACATCAGG
	4001	ACGAATATTG	TCCCCTGGCT	ATCCAGCTCC	GTATGACAAC	AACCTCCACT
	4051	GCACCTGGAT	TATAGAGGCA	GACCCAGGAA	AGACCATTAG	CCTCCATTTC
	4101	ATTGTTTTCG	ACACGGAGAT	GGCTCACGAC	ATCCTCAAGG	TCTGGGACGG
	4151	GCCGGTGGAC	AGTGACATCC	TGCTGAAGGA	GTGGAGTGGC	TCCGCCCTTC
25	4201	CGGAGGACAT	CCACAGCACC	TTCAACTCAC	TCACCCTGCA	GTTCGACAGC
•	4251	GACTTCTTCA	TCAGCAAGTC	TGGCTTCTCC	ATCCAGTTCT	CCACCTCAAT
	4301	TGCAGCCACC	TGTAACGATC	CAGGTATGCC	CCAAAATGGC	ACCCGCTATG
	4351	GAGACAGCAG	AGAGGCTGGA	GACACCGTCA	CATTCCAGTG	TGACCCTGGC
	4401	TATCAGCTCC	AAGGACAAGC	CAAAATCACC	TGTGTGCAGC	TGAATAACCG
30	4451	GTTCTTTTGG	CAACCAGACC	CTCCTACATG	CATAGCTGCT	TGTGGAGGGA
	4501	ATCTGACGGG	CCCAGCAGGT	GTTATTTTGT	CACCCAACTA	CCCACAGCCG
	4551	TATCCTCCTG	GGAAGGAATG	TGACTGGAGA	GTAAAAGTGA	ACCCGGACTT
	4601	TGTCATCGCC	TTGATATTCA	AAAGTTTCAA	CATGGAGCCC	AGCTATGACT
	4651	TCCTACACAT	CTATGAAGGG	GAAGATTCCA	ACAGCCCCCT	CATTGGGAGT
35	4701	TACCAGGGCT	CTCAGGCCCC	AGAAAGAATA	GAGAGTAGCG	CCCMMCCCCM
•	4751	GTTTCTGGCA	TTTCGGAGTG	ATGCCTCCGT	GGGCCTTTCA	DCCDDAMAMA
	4801	TTGAATTTAA	AGAGAAACCA	CGGGAAGCTT	BACCEE	AGGAAATATA
	4851	ATGAATGGGA	CAAGAGTTGG	AACAGACTTC	AAGCTTGGCT	CCACCATCAC
	4901	CTACCAGTGT	GACTCTGGCT	ATAAGATTCT	CCCACCAACT	CCTCCCCTCC
40	4951	GTGTGATTGG	GGCTGATGGG	AAACCCTCCT	CCARCACAAGI	CCCTACTTT
	5001	TGCAATGCTC	CCTGTGGAGG	CCAGTACACG	TCCTCA A A TA	TCCCTCTATT
	5051	ATCACCAAAC	TACCCCCATA	ATTACACAGC	TGGICAAAIA	TGCCTCTATT
	5101	CCATCACGGT	ACCAAAGGAA	TTCGTGGTCT	TIGGACAGII	CCCATGCACA
	5151	CAGACAGCCC	TGAATGATTT	GGCAGAATTA	TITORIGOM	CANACATTCC
45	5201	GGCCAGACTT	CTCAGCTCAC	TCTCGGGGTC	CATTCAGGG	DANACATIC
	5251	CCTTGGCTAC	GTCAAATCAA	ATTCTGCTCC	CALICAGIGC	CTCCTACCAC
	5301	GCCTCTGCCC	GCGGCTTCCA	CTTCGTGTAT	CAAGCIGIIC	ACCACAATTC
•	5351	TGACACCCAA	TGCAGCTCTG	TCCCCGAGCC	CAGAIACGGA	CAACCCGGA
	5401	GTTCTGAGTT	TTCTGCCGGC	TCCATCGTCC	MCCCACTC	TCCCCD ACCC
50	5451	TACCTGCTTC	AGGGTTCCAC	GGCGCTCCAC	TGCCAGTCCG	CCCTCCAACGC
	5501	CTTGGCACAG	TGGAACGACA	CGATCCCCAG	CIGIGIGGIA	CTACCCTCAC
	5551	GCAATTTCAC	TCAACGAAGA	GGTACAATCC	TGTCCCCCGG	DDD CCC DCCC
	5601	CCATACGGAA	ACAACTTGAA	CTGTATATGG	MAGATCATAG	CACAACTCC
	5651	CTCGGGAATT	CAGATCCAAG	TGATCAGTTT	TGCCACGGAG	CAGAACIGGG
55		ACTCCCTTGA	GATCCACGAT	GGTGGGGATG	TGACCGCACC	CHONCINGON
	5751	AGCTTCTCAG	GCACCACAGT	ACCGGCACTG	TGAACAGTA	CTTCCAACCA
	5801	ACTCTACCTG	CATTTCCAGT	CTGACATTAG	CARCCARCT	GCTGGTTTCC
	5851	ACCTGGAATA	CAAAACTGTA	GGTCTTGCTG	CATGCCAAGA	ACCAGCCCTC
	5901	CCCAGCAACA	GCATCAAAAT	CGGAGATCGG	TACATGGTGA	ACGACGTGCT
60	5951	CTCCTTCCAG	TGCGAGCCCG	GGTACACCCT	ACMARCCCEC	TCCCACATTT
	6001	CCTGTATGCC	AGGGACCGTT	CGCCGTTGGA	MULCOMOCONO	TCCCCTGTGC
	6051	ATTGCAACCT	GTGGAGGGAC	GCTGAGCACC	COUNCICO	TGATCCTGAG
	6101	CCCCGGCTTC	CCAGGTTCTT	ACCCCAACAA	CTTAGACTGC	ACCTGGAGGA

		6151	TCTCATTACC	CATCGGCTAT	GGTGCACATA	TTCAGTTTCT	GAATTTTTCT
		6201	ACCGAAGCTA		CCTTGAAATT	CAAAATGGAC	CTTACCACAC
1 222		6251	CAGCCCCATG	ATTGGACAAT	TTAGCGGCAC	GGATCTCCCC	GCGGCCCTGC
. ~~	;	6301	TGAGCACAAC		CTCATCCACT	TTTATAGTGA	CCATTCGCAA
5		6351	AACCGGCAAG	GATTTAAACT	TGCTTACCAA	GCCTATGAAT	TACAGAACTG
•		6401	TCCAGATCCA	CCCCCATTTC	AGAATGGGTA	CATGATCAAC	TCGGATTACA
		6451	GCGTGGGGCA	ATCAGTATCT	TTCGAGTGTT	ATCCTGGGTA	CATTCTAATA
		6501	GGCCATCCTG	TCCTCACTTG	TCAGCATGGG	ATCAACAGAA	ACTGGAACTA
٠		6551	CCCTTTTCCA	AGATGTGATG	CCCCTTGTGG	GTACAACGTA	ACTTCTCAGA
1.0.	•	6601	ACGGCACCAT	CTACTCCCCT	GGCTTTCCTG	ATGAGTATCC	GATCCTGAAG
٠	•	6651	GACTGCATTT	GGCTCATCAC	GGTGCCTCCA	GGGCACGGAG	TTTACATCAA
		6701	CTTCACCCTG	TTACAGACGG	AAGCTGTCAA	CGATTACATT	GCTGTTTGGG
		6751	ACGGTCCCGA	TCAGAACTCA	CCCCAGCTGG	GAGTTTTCAG	TGGCAACACA
		6801	GCCCTCGAAA	ÇÇÇÇTATAÇ	CTCCACCAAC	CAAGTCCTGC	TCAAGTTCCA
15 .		6851	CAGCGACTTT	TCAAATGGAG	GCTTCTTTGT	CCTCAATTTC	CACGGTCAGT
	•	6901	TGATTTTCAC	TCCGTTAGTT	AAGACTGAGA	ATTCCATGTG	GTGTTTACTG
	•	6951	CAGTGTTGTC	CCACGCCTTG	TTTCCAGCTG	AAGTTTCTTG	ATTCAGCCGA
		7001	GGGCGTGTAT	GATTCTTTTG	CACTGGAGGC	CAGCGTTTCC	TGTGGTCCTT
		7051	TTTTTGTTTA	ATGATGTCTT	TATTATTTCA	CATCGTATCC	AGCTTGGATT
20		7101	TATTCCAAGA	TACATGTATC	CTAAGTGAAA	CTCTAAGATG	AAGACCATTG
		7151	AAAGAGATTT	GGTACCTTTT	ATAGATTTAC	TCATCCCTGT	CTCAAGATAA
		7201	GGTGTTATAG	CAAATGTCAT	GTAACTATAA	ATGGTGTGAA	AGCAAACCTC
		7251	CAATAATCCT	GGGAATGCAC	TCTAAACGAT	ATGTAGAACA	TCTGTCAATC
		7301	NATCGCTTAT	CTCTCACGAA	CAC		

SEQ ID NO:6

5R23V2

20 CCAATGGCACTATTGAGAGCCCAGGGTTTCCTCACGGGTATCCGAACTATGCCAACTGCA CCTGGATCATCACGGGCGAGCGCAATAGGATACAGTTGTCCTTCCATACCTTTGCTCT TGAAGAAGATTTTGATATTTTATCAGTTTACGATGGACAGCCTCAACAAGGGAATTTAAA AGTGAGATTATCGGGATTTCAGCTGCCCTCCTCTATAGTGAGTACAGGATCTATCCTCACT CTGTGGTTCACGACAGACTTCGCTGTGAGTGCCCAAGGTTTCAAAGCATTATATGAAGTT

TTACCTAGCCACACTTGTGGAAATCCTGGAGAAATCCTGAAAGGAGTTCTGCATGGAACG
AGATTCAACATAGGAGACAANATCCGGTACAGCTGCCTCCCTGGCTACATCTTGGAAGGC
CACGCCATCCTGCATCGTCAGCCCAGGAAATGGTGCATCGTGGGACTTCCCAGCT
CCCTTTTGCAGAGCTGAGGGAGCCTGCGGAGGAACCTTACGCGGGACCAGCAGCTCCATC
TCCAGCCCGCACTTCCCTTCAGAGTACGAGAACAACGCGGACTGCACCTGGACCATTCTG

30 GCTGAGCCCGGGGACACCATTGCGCTGGTCTTCACTGACTTTCAGCTAGAAGAAGAAGATAT GATTTCTTAGAGATCAGTGGCACGGAAGCTCCATCCATATGGCTAACTGGCATGAACCTC CCCTCTCCAGTTATCAGTAGCAAGAATTGGCTACGACTCCATTTCACCTCTGACAGCAACC ACCGACGCAAAGGATTTAACGCTCAGTTCCAAGTGAAAAAGGCGATTGAGTTGAAGTCA AGAGGAGTCAAGATGCCCCAGCAAGGATGGAAGCCATAAAAACTCTGTCTTGAGCCA

40 GGTCATCACCACCACCGGACAAGGTCATCAAGCTTGCCTTNGAAGAGTTTGAGCT
GGAGCGAGGCTATGACACCCTNACGGTTGGTGATGCTGGGAAGGTGGGAGACACCAGAT
CGGTCTTGTANGTGCTCACGGGATCCAGTGTTCCTGACCTCATTGTGAGCATGAGCAACC
AGATGTGGCTACATCTGCAGTCGGATGATAGCATTGGCTCACCTGGGTTTAAAGCTGTTT
ACCAAGAAATTGAAAAGGGAGGGTGTGGGGGATCCTGGAATCCCCGCCTATGGGAAGCGG

45 ACGGGCAGCAGTTTCCTCCATGGAGATNCACTNACCTTTGAATGCCCGGCGGCCTTTGAG
CTGGTGGGGGAGAGAGTTATCACCTGTCAGCAGAACAATCAGTGGTCTGGCAACAAGCCC
AGCTGTGTATTTTCATGTTTCTTCAACTTTACGGCATCATCTGGGATTATTCTGTCACCAA
ATTATCCAGAGGAATATGGGAACAACATGAACTGTGTCTGGTTGATTATCTCGGAGCCAG
GAAGTCGAATTCACCTAATCTTTAATGATTTTGATGTTGAGCCTCAATTTGACTTTCTCGC

50 GGTCAAGGATGACCTAATCTTTAATGATTTCATCTGCACCTTTTCTGGCAATGAAGT
GCCTTCCCAGCTGGCCAGCAGTGGGCATATAGTTCGCTTGGAATTTCAGTCTGACCATTCC
ACTACTGGCAGAGGGTTNAACATCACTTACACCACNTTTGGTCAGAATGAGTGCCATGAT
CCTGGCATTCCTATAAACGGACGACGTTTTGGTGACAGGTTTCTACTCGGGAGCTCGGTT
TCTTTCCACTGTGATGATGGCTTTGTCAAGACCCAGGGATCCGAGTCCATTACCTGCATAC
55 TGCAAGACGGGAACGTGGTCTGGAGCTCCACCGTGCCCCGCTGTGAAGCTCCATGTGGTG

GTTCGTCCCACTGATCGCGAGTACCACGGCACCCAGGCACCCCAGTTCCTCATCAGCA CCGGGAACTTCATGTACCTGCTATTCACCACTGACAACAGCCGCTCCAGCATCGGCTTCCT ::::: 5 GAACGGCCATCGCCACGGTGGAGACTTTGGCATCAGGTCCACAGTGACTTTCAGCTGTGA CCCGGGGTACACACTAAGTGACGACGAGCCCCTCGTCTGTGAGAGGAACCACCAGTGGA ACCACGCCTTGCCCAGCTGCGACGCTCTATGTGGAGGCTACATCCAAGGGAAGAGTGGAA CAGTCCTTTCTCCTGGGTTTCCAGATTTTTATCCAAACTCTCTAAACTGCACGTGGACCAT TGAAGTGTCTCATGGGAAAGGAGTTCAAATGATCTTTCACACCTTTCATCTTGAGAGTTCC 10 CACGACTATTTACTGATCACAGAGGATGGAAGTTTTTCCGAGCCCGTTGCCAGGCTCACC GGGTCGGTGTTGCCTCATACGATCAAGGCAGGCCTGTTNGGAAACTTCACTGCCCAGCTT CGGTTTATATCAGACTTCTCAATTTCGTACGAGGGCTTCAATATCACATTTTCAGAATATG ACCTGGAGCCATGTGATGATCCTGGAGTCCCTGCCTTCAGCCGAAGAATTGGTTTTCACTT TGGTGTGGGAGACTCTCTGACGTTTTCCTGCTTCCTGGGATATCGTTTAGAAGGTGCCACC 15 AAGCTTACCTGCCTGGGTGGGGCCGCCGTGTGTGGAGTGCACCTCTGCCAAGGTGTGTG GCCGAATGTGGAGCAAGTGTCAAAGGAAATGAAGGAACATTACTGTCTCCAAATTTTCCA TCCAATTATGATAATAACCATGAGTGTATCTATAAAATAGAAACAGAAGCCGGCAAGGGC ATCCACCTTAGAACACGAAGCTTCCAGCTGTTTGAAGGAGATACTCTAAAGGTATATGAT GGAAAAGACAGTTCCTCACGTCCACTGGGCACGTTCACTAAAAATGAACTTCTGGGGCTG 20 ATCCTAAACAGCACATCCAATCACCTGTGGCTAGAGTTCAACACCAATGGATCTGACACC GACCAAGGTTTTCAACTCACCTATACCAGTTTTGATCTGGTAAAATGTGAGGATCCGGGC ATCCCTAACTACGGCTATAGGATCCGTGATGAAGGCCACTTTACCGACACTGTAGTTCTG TACAGTTGCAACCCGGGGTACGCCATGCATGGCAGCAACACCCTGACCTGTTTGAGTGGA GACAGGAGAGTGTGGGACAAACCACTACCTTCGTGCATAGCGGAATGTGGTGGTCAGAT 25 CCATGCAGCCACATCAGGACGAATATTGTCCCCTGGCTATCCAGCTCCGTATGACAACAA CCTCCACTGCACCTGGATTATAGAGGCAGACCCAGGAAAGACCATTAGCCTCCATTTCAT TGTTTTCGACACGGAGATGGCTCACGACATCCTCAAGGTCTGGGACGGGCCGGTGGACAG TGACATCCTGCTGAAGGAGTGGAGTGGCTCCGCCCTTCCGGAGGACATCCACAGCACCTT CAACTCACTCACCTGCAGTTCGACAGCGACTTCTTCATCAGCAAGTCTGGCTTCTCCATC 30 CAGTTCTCCACCTCAATTGCAGCCACCTGTAACGATCCAGGTATGCCCCAAAATGGCACC CGCTATGGAGACAGCAGAGAGGCTGGAGACACCGTCACATTCCAGTGTGACCCTGGCTAT CAGCTCCAAGGACAAGCCAAAATCACCTGTGTGCAGCTGAATAACCGGTTCTTTTGGCAA CCAGACCCTCCTACATGCATAGCTGCTTGTGGAGGGAATCTGACGGGCCCAGCAGGTGTT 35 AAAGTGAACCCGGACTTTGTCATCGCCTTGATATTCAAAAGTTTCAACATGGAGCCCAGC TATGACTTCCTACACATCTATGAAGGGGAAGATTCCAACAGCCCCCTCATTGGGAGTTAC CAGGGCTCTCAGGCCCCAGAAAGAATAGAGAGTAGCGGAAACAGCCTGTTTCTGGCATTT CGGAGTGATGCCTCCGTGGGCCTTTCAGGGTTCGCCATTGAATTTAAAGAGAAACCACGG GAAGCTTGTTTTGACCCAGGAAATATAATGAATGGGACAAGAGTTGGAACAGACTTCAAG 40 CTTGGCTCCACCATCACCTACCAGTGTGACTCTGGCTATAAGATTCTTGACCCCTCATCCA TCACCTGTGTGATTGGGGCTGATGGGAAACCCTCCTGGGACCAAGTGCTGCCCTCCTGCA ATGCTCCCTGTGGAGGCCAGTACACGGGATCAGAAGGGGTAGTTTTATCACCAAACTACC CCCATAATTACACAGCTGGTCAAATATGCCTCTATTCCATCACGGTACCAAAGGAATTCG TGGTCTTTGGACAGTTTGCCTATTTCCAGACAGCCCTGAATGATTTGGCAGAATTATTTGA 45 ATTGCCCTTGGCTACGTCAAATCAAATTCTGCTCCGATTCAGTGCAAAGAGCGGTGCCTCT GCCGCGGCTTCCACTTCGTGTATCAAGCTGTTCCTCGTACCAGTGACACCCAATGCAGCT CTGTCCCGAGCCCAGATACGGAAGGAGAATTGGTTCTGAGTTTTCTGCCGGCTCCATCG TCCGATTCGAGTGCAACCCGGGATACCTGCTTCAGGGTTCCACGGCGCTCCACTGCCAGT 50 CCGTGCCCAACGCCTTGGCACAGTGGAACGACACGATCCCCAGCTGTGTGGTACCCTGCA GTGGCAATTTCACTCAACGAAGAGGTACAATCCTGTCCCCCGGCTACCCTGAGCCATACG GAAACAACTTGAACTGTATATGGAAGATCATAGTTACGGAGGGCTCGGGAATTCAGATCC AAGTGATCAGTTTTGCCACGGAGCAGAACTGGGACTCCCTTGAGATCCACGATGGTGGGG ATGTGACCGCACCCAGACTGGGAAGCTTCTCAGGCACCACAGTACCGGCACTGCTGAACA 55 GTACTTCCAACCAACTCTACCTGCATTTCCAGTCTGACATTAGTGTGGCAGCTGCTGGTTT CCACCTGGAATACAAAACTGTAGGTCTTGCTGCATGCCAAGAACCAGCCCTCCCCAGCAA

CAGCATCAAAATCGGAGATCGGTACATGGTGAACGACGTGCTCTCCTTCCAGTGCGAGCC

CGGGTACACCCTGCAGGGCCGTTCCCACATTTCCTGTATGCCAGGGACCGTTCGCCGTTG
GAACTATCCGTCTCCCCTGTGCATTGCAACCTGTGGAGGGACGCTGAGCACCTTGGGTGG
TGTGATCCTGAGCCCCGGCTTCCCAGGTTCTTACCCCAACAACTTAGACTGCACCTGGAG
GATCTCATTACCCATCGGCTATGGTGCACAACTTACCAACACTTCGATTGGACAA

5 AATCATGACTTCCTTGAAATTCAAAATGGACCTTACCACACCAGCCCCATGATTGGACAA
TTTAGCGGCACGGATCTCCCCGCGGCCCTGCTGAGCACAACGCATGAAACCCTCATCCAC
TTTTATAGTGACCATTCGCAAAACCGGCAAGGATTTAAACTTGCTTACCAAGCCTATGAA
TTACAGAACTGTCCAGATCCACCCCCATTTCAGAATGGGTACATGATCAACTCGGATTAC
AGCGTGGGGCAATCAGTATCTTTCGAGTGTTATCCTGGGTACATTCTAATAGGCCATCCT

10" GTCCTCACTTGTCAGCATGGGATCAACAGAÄACTGGAACTACCCTTTTCCAAGATGTGAT GCCCCTTGTGGGTACAACGTAACTTCTCAGAACGGCACCATCTACTCCCCTGGCTTTCCTG ATGAGTATCCGATCCTGAAGGACTGCATTTGGCTCATCACGGTGCCTCCAGGGCACGGAG TTTACATCAACTTCACCCTGTTACAGACGGAAGCTGTCAACGATTACATTGCTGTTTTGGGA CGGTCCCGATCAGAACTCACCCCAGCTGGGAGTTTTCAGTGGCAACACACAGCCCTCGAAAC

20 GTATCCTAAGTGAAACTCTAAGATGAAGACCATTGAAAGAGATTTGGTACCTTTTATAGA TTTACTCATCCCTGTCTCAAGATAAGGTGTTATAGCAAATGTCATGTAACTATAAATGGTG TGAAAGCAAACCTCCAATAATCCTGGGAATGCACTCTAAACGATATGTAGAACATCTGTC AATCNATCGCTTATCTCTCACGAACACN

SEQ ID NO:7

5R2 OC147

5 AAGGGAGATCAAAGGATGGAGGCGGGACTCTGCCCCTGCAGAAACCCTCCAGTTTGCTGGAGTTGCCG GATTACATTGTTCCTCCCGGTGTGCGGCGTGAGCTTCCCCCACCGAGCGCCCCAACAAGTCTCCTTT CTCCAGCCTGCGCGCTGCTGCGCTGAGGCCGAATGAAGCGCAGCACGGTGCGGGCAGCCCGAGGCCCC GAGGCTGGGCTCTGTCTGTGGGACTGCGCCGTGCCCAGCCTCGGTCCCCTCTCTGTGGGTAAGGAT GGTTGAGTCCAGCCTCCACGGCAGCGGCTCCTTGTGCCACTAGCAGCCCTTCTTCTGCGCTCTCCGCC 10 CGCAGCCTCGČCGCCTTGGTGCCTTCCTGCCCGGCTCGGCCGCCTCGTCCCCGGCCCCGGCCCCGC CAGCCCGGGTCTCCGCGCTCGGAGCAGCTCAGCCCTGCAGTGGCTCGGGACCCGATGCTATGAGAGGG AAGCGAGCCGGGCGCCCAGACCTTCAGGAGGCGTCGGATGCGCGGGGTCTTGGGACCGGGCTCTCT 15 CTCCGGCTCGCCTTGCCCTCGGGTGATTATTTGGCTCCGCTCATAGCCCTGCCTTCCTCGGAGGAGCC ATCGGTGTCGCGTGCGTGTGGAGTATCTGCAGACATGACTGCGTGGAGGAGATTCCAGTCGCTGCTCC TGCTTCTCGGGCTGCTGGTGCTGCGCGAGGCTCCTCACTGCAGCGAAGGGTCAGAACTGTGGAGGC TTAGTCCAGGGTCCCAATGGCACTATTGAGAGCCCAGGGTTTCCTCACGGGTATCCGAACTATGCCAA CTGCACCTGGATCATCACGGGCGAGCGCAATAGGATACAGTTGTCCTTCCATACCTTTGCTCTTG 20 AAGAAGATTTTGATATTTTATCAGTTTACGATGGACAGCCTCAACAAGGGAATTTAAAAAGTGAGATTA TCGGGATTTCAGCTGCCCTCCTCTATAGTGAGTACAGGATCTATCCTCACTCTGTGGTTCACGACAGA CTTCGCTGTGAGTGCCCAAGGTTTCAAAGCATTATATGAAGTTTTACCTAGCCACACTTGTGGAAATC CTGGAGAAATCCTGAAAGGAGTTCTGCATGGAACGAGATTCAACATAGGAGACAAAATCCGGTACAGC TGCCTCCCTGGCTACATCTTGGAAGGCCACGCCATCCTGACCTGCATCGTCAGCCCAGGAAATGGTGC 25 ATCGTGGGACTTCCCAGCTCCCTTTTGCAGAGCTGAGGGAGCCTGCGGAGGAACCTTACGCGGGACCA GCAGCTCCATCTCCAGCCCGCACTTCCCTTCAGAGTACGAGAACAACGCGGACTGCACCTGGACCATT CTGGCTGAGCCCGGGGACACCATTGCGCTGGTCTTCACTGACTTTCAGCTAGAAGAAGGATATGATTT CTTAGAGATCAGTGGCACGGAAGCTCCATCCATATGGCTAACTGGCATGAACCTCCCCTCTCCAGTTA TCAGTAGCAAGAATTGGCTACGACTCCATTTCACCTCTGACAGCAACCACCGACGCAAAGGATTTAAC 30 GCTCAGTTCCAAGTGAAAAAGGCGATTGAGTTGAAGTCAAGAGGAGTCAAGATGCTGCCCAGCAAGGA TGGAAGCCATAAAAACTCTGTCTGTGAGTCCCTTTCCTTTCTATCTGAGGATTGATACGCCCTTGTAA GCAGAGGAGAATGGAGCAGTG

5R2 AW

AAGGGAGATCAAAGGATGGAGGCGGGACTCTGCCCCTGCAGAAACCCTCCAGTTTGCTGGAGTTGCCG 5 GATTACATTGTTCCTCCCCGGTGTGCGGCGTGAGCTTCCCCCACCCGAGCGCCCAACAAGTCTCCTTT CTCCAGCCTGCGCGCTGCGCCTGAGGCCGAATGAAGCGCAGCACGGTGCGGGCAGCCCGAGGCCCC GAGGCTGGGCTCTGTCTGGGACTGCGCCGTGCCCAGCCTCGGTCCCCTCTCTGTGGGTAAGGAT GGTTGAGTCCAGCCTCCACGGCAGCGGCTCCTTGTGCCACTAGCAGCCCTTCTTCTGCGCTCTCCGCC 10 CAGCCCGGGTCTCCGCGCTCGGAGCAGCTCAGCCCTGCAGTGGCTCGGGACCCGATGCTATGAGAGGG AAGCGAGCCGGGCGCCCAGACCTTCAGGAGGCGTCGGATGCGCGGGGGGTCTTGGGACCGGGCTCTCT CTCCGGCTCGCCTTGCCCTCGGGTGATTATTTGGCTCCGCTCATAGCCCTGCCTTCCTCGGAGGAGCC 15 ATCGGTGTCGCGTGCGTGTGGAGTATCTGCAGACATGACTGCGTGGAGGAGATTCCAGTCGCTGCTCC TGCTTCTCGGGCTGCTGGTGCTGTGCGCGAGGCTCCTCACTGCAGCGAAGGGTCAGAACTGTGGAGGC TTAGTCCAGGGTCCCAATGGCACTATTGAGAGCCCAGGGTTTCCTCACGGGTATCCGAACTATGCCAA CTGCACCTGGATCATCACGGGCGAGCGCAATAGGATACAGTTGTCCTTCCATACCTTTGCTCTTG AAGAAGATTTTGATATTTTATCAGTTTACGATGGACAGCCTCAACAAGGGAATTTAAAAGTGAGATTA 20 TCGGGATTTCAGCTGCCCTCCTCTATAGTGAGTACAGGATCTATCCTCACTCTGTGGTTCACGACAGA CTTCGCTGTGAGTGCCCAAGGTTTCAAAGCATTATATGAAGTTTTACCTAGCCACACTTGTGGAAATC CTGGAGAAATCCTGAAAGGAGTTCTGCATGGAACGAGATTCAACATAGGAGACAAAATCCGGTACAGC TGCCTCCCTGGCTACATCTTGGAAGGCCACGCCATCCTGACCTGCATCGTCAGCCCAGGAAATGGTGC ATCGTGGGACTTCCCAGCTCCCTTTTGCAGAGCTGAGGGAGCCTGCGGAGGAACCTTACGCGGGACCA 25 GCAGCTCCATCTCCAGCCCGCACTTCCCTTCAGAGTACGAGAACAACGCGGACTGCACCTGGACCATT CTGGCTGAGCCCGGGGACACCATTGCGCTGGTCTTCACTGACTTCAGCTAGAAGAAGGATATGATTT CTTAGAGATCAGTGGCACGGAAGCTCCATCCATATGGCTAACTGGCATGAACCTCCCCTCTCCAGTTA TCAGTAGCAAGAATTGGCTACGACTCCATTTCACCTCTGACAGCAACCACCGACGCAAAGGATTTAAC GCTCAGTTCCAAGTGAAAAAGGCGATTGAGTTGAAGTCAAGAGGAGTCAAGATGCTGCCCAGCAAGGA 30 TCATGACAAGGAATGGGAGAATTTCCCTGACAGCCTCAGGAAACTTGCAGTTTGATAATTAAACAGAT CAAGGTCACTCAGATGAGCTGATGGGACATGCTGTGTACGGAGGAGCATTTGCAGTTACAACACTTTG TAGCCATGCAGGATGGGGCAATTAATCCAGAACCATTATTTAATAAAAAGATGATTTTTTAAATGTGA 35

protein sequence >ORF:121..5598 Frame +1

.5

MEAIKTLSGIWNNINHVTSEEDTFIMYLGKPWLQVKIQVSQGGVALVSDMCPDPGIPENGRRAGSDFR VGANVQFSCEDNYVLQGSKSITCQRVTETLAAWSDHRPICRARTCGSNLRGPSGVITSPNYPVQYEDN AHCVWVITTTDPDKVIKLAFEEFELERGYDTLTVGDAGKVGDTRSVLYVLTGSSVPDLIVSMSNOMWL 10 HLQSDDSIGSPGFKAVYQEIEKGGCGDPGIPAYGKRTGSSFLHGDTLTFECPAAFELVGERVITCQQN NOWSGNKPSCVFSCFFNFTASSGIILSPNYPEEYGNNMNCVWLIISEPGSRIHLIFNDFDVEPQFDFL AVKDDGISDITVLGTFSGNEVPSQLASSGHIVRLEFQSDHSTTGRGFNITYTTFGQNECHDPGIPING RRFGDRFLLGSSVSFHCDDGFVKTQGSESITCILQDGNVVWSSTVPRCEAPCGGHLTASSGVILFPGW PGYYKDSLHCEWIIEAKPGHSIKITFDRFQTEVNYDTLEVRDGPASSSPLIGEYHGTQAPQFLISTGN 15 FMYLLFTTDNSRSSIGFLIHYESVTLESDSCLDPGIPVNGHRHGGDFGIRSTVTFSCDPGYTLSDDEP LVCERNHQWNHALPSCDALCGGYIQGKSGTVLSPGFPDFYPNSLNCTWTIEVSHGKGVQMIFHTFHLE SSHDYLLITEDGSFSEPVARLTGSVLPHTIKAGLFGNFTAQLRFISDFSISYEGFNITFSEYDLEPCD DPGVPAFSRRIGFHFGVGDSLTFSCFLGYRLEGATKLTCLGGGRRVWSAPLPRCVAECGASVKGNEGT LLSPNFPSNYDNNHECIYKIETEAGKGIHLRTRSFQLFEGDTLKVYDGKDSSSRPLGTFTKNELLGLI 20 LNSTSNHLWLEFNTNGSDTDQGFQLTYTSFDLVKCEDPGIPNYGYRIRDEGHFTDTVVLYSCNPGYAM HGSNTLTCLSGDRRVWDKPLPSCIAECGGQIHAATSGRILSPGYPAPYDNNLHCTWI1EADPGKTISL HFIVFDTEMAHDILKVWDGPVDSDILLKEWSGSALPEDIHSTFNSLTLQFDSDFFISKSGFSIOFSTS IAATCNDPGMPONGTRYGDSREAGDTVTFQCDPGYQLQGQAKITCVQLNNRFFWQPDPPTCIAACGGN LTGPAGVILSPNYPQPYPPGKECDWRVKVNPDFVIALIFKSFNMEPSYDFLHIYEGEDSNSPLIGSYQ GSQAPERIESSGNSLFLAFRSDASVGLSGFAIEFKEKPREACFDPGNIMNGTRVGTDFKLGSTITYOC 25 DSGYKILDPSSITCVIGADGKPSWDOVLPSCNAPCGGQYTGSEGVVLSPNYPHNYTAGQICLYSITVP KEFVVFGQFAYFQTALNDLAELFDGTHAQARLLSSLSGSHSGETLPLATSNQILLRFSAKSGASARGF HFVYQAVPRTSDTQCSSVPEPRYGRRIGSEFSAGSIVRFECNPGYLLQGSTALHCQSVPNALAQWNDT IPSCVVPCSGNFTQRRGTILSPGYPEPYGNNLNCIWKIIVTEGSGIQIQVISFATEQNWDSLEIHDGG 30 DVTAPRLGSFSGTTVPALLNSTSNQLYLHFQSDISVAAAGFHLEYKTVGLAACQEPALPSNSIKIGDR YMVNDVLSFOCEPGYTLQGRSHISCMPGTVRRWNYPSPLCIATCGGTLSTLGGVILSPGFPGSYPNNL . DCTWRISLPIGYGAHIOFLNFSTEANHDFLEIQNGPYHTSPMIGQFSGTDLPAALLSTTHETLIHFYS DHSQNRQGFKLAYQAYELQNCPDPPPFQNGYMINSDYSVGQSVSFECYPGYILIGHPP

:	50-3V1 Pro	otein sequen	ice	1801 AA		
	1	MEAIKTLSGI	WNNINHVTSE	EDTFIMYLGK	PWLQVKIQVS	QGGVALVSDM
5	51		RRAGSDFRVG	ANVQFSCEDN	YVLQGSKSIT	CQRVTETLAA
٠	101	WSDHRPICRA	· · · · · · · · · · · · · · · · · · ·	SGVITSPNYP	VQYEDNAHCV	WVITTTDPDK
	151	VIKLAFEEFE.		GDAGKVGDTR	SVLYVLTGSS	VPDLIVSMSN
	201	OMWLHLOSDD	SIGSPGFKAV	YQEIEKGGCG	DPGIPAYGKR	TGSSFLHGDT
٠,	, 251	LTFECPAAFE	LVGERVITCQ	QNNQWSGNKP	SCVFSCFFNF	TASSGIILSP
10	301	NYPEEYGNNM	NCVWLIISEP	GSRIHLIFND	FDVEPQFDFL	AVKDDGISDI
·	351	TVLGTFSGNE	VPSQLASSGH	IVRLEFQSDH	STTGRGFNIT	YTTFGQNECH
	401	DPGIPINGRR	FGDRFLLGSS	VSFHCDDGFV	KTQGSESITC	ILQDGNVVWS
	451	STVPRCEAPC	GGHLTASSGV	ILPPGWPGYY	KDSLHCEWII	EAKPGHSIKI
	501	TFDRFQTEVN	YDTLEVRDGP	ASSSPLIGEY	HGTQAPQFLI'	
15	551	TTDNSRSSIG	FLIHYESVTL	ESDSCLDPGI	PVNGHRHGGD	FGIRSTVTFS
	601	CDPGYTLSDD	EPLVCERNHQ	WNHALPSCDA	LCGGYIQGKS	GTVLSPGFPD
	651	FYPNSLNCTW	TIEVSHGKGV	QMIFHTFHLE	SSHDYLLITE	DGSFSEPVAR
	701	LTGSVLPHTI	KAGLFGNFTA	QLRFISDFSI	SYEGFNITFS	EYDLEPCDDP
•	751	GVPAFSRRIG	FHFGVGDSLT	FSCFLGYRLE	GATKLTCLGG	GRRVWSAPLP
20	801	RCVAECGASV	KGNEGTLLSP	NFPSNYDNNH	ECIYKIETEA	GKGIHLRTRS
	851	FQLFEGDTLK	VYDGKDSSSR	PLGTFTKNEL	LGLILNSTSN	HLWLEFNTNG
****	901	SDTDQGFQLT	YTSFDLVKCE	DPGIPNYGYR	IRDEGHFTDT	VVLYSCNPGY
	951	AMHGSNTLTC	LSGDRRVWDK	PLPSCIAECG	GQIHAATSGR	ILSPGYPAPY
	1001	DNNLHCTWII	EADPGKTISL	HFIVFDTEMA	HDILKVWDGP	VDSDILLKEW
25	1051	SGSALPEDIH	STFNSLTLQF	DSDFFISKSG	FSIQFSTSIA	
	1101	NGTRYGDSRE	AGDTVTFQCD	PGYQLQGQAK	ITCVQLNNRF	FWQPDPPTCI
	1151	AACGGNLTGP	AGVILSPNYP	QPYPPGKECD	WRVKVNPDFV	IALIFKSFNM
	. 1201	EPSYDFLHIY	EGEDSNSPLI	GSYQGSQAPE	RIESSGNSLF	LAFRSDASVG
	1251	LSGFAIEFKE	KPREACFDPG	NIMNGTRVGT	DFKLGSTITY	QCDSGYKILD
30	1301	PSSITCVIGA	DGKPSWDQVL	PSCNAPCGGQ	YTGSEGVVLS	PNYPHNYTAG RLLSSLSGSH
	1351	QICLYSITVP	KEFVVFGQFA	YFOTALNDLA	ELFDGTHAQA	
	1401	SGETLPLATS	NQILLRFSAK		VYQAVPRTSD	TQCSSVPEPR AOWNDTIPSC
	1451	YGRRIGSEFS	AGSIVRFECN	PGYLLQGSTA	LHCQSVPNAL IWKIIVTEGS	GIOIOVISFA
	1501	VVPCSGNFTQ	RRGTILSPGY	PEPYGNNLNC	ALLNSTSNQL	YLHFQSDISV
35	1551	TEQNWDSLEI	HDGGDVTAPR		DRYMVNDVLS	FOCEPGYTLO
	1601	AAAGFHLEYK		ALPSNSIKIG	STLGGVILSP	GFPGSYPNNL
	1651	GRSHISCMPG	TVRRWNYPSP	LCIATCGGTL	-	PMIGOFSGTD
	1701	DCTWRISLPI	GYGAHIQFLN	FSTEANHDFL		KPKSKYTSYM
	1751	LPAALLSTTH	ETLIHFYSDH	20NKOGE KTW	YQGMEQQREP	KIKSKIISIM
40	1801	*				

	5G-3V2 Protein sequence		2009 AA			
	1	MEAIKTLSGI	WNNINHVTSE	EDTFIMYLGK	PWLQVKIQVS	QGGVALVSDM
5	· 51	CPDPGIPENG	RRAGSDFRVG	ANVQFSCEDN	YVLQGSKSIT	CQRVTETLAA
	101	WSDHRPICRA	RTCGSNLRGP	SGVITSPNYP	VQYEDNAHCV	WVITTTDPDK
	151	VIKLAFEEFE	LERGYDTLTV	GDAGKVGDTR	SVLYVLTGSS	VPDLIVSMSN
	201	QMWLHLQSDD	SIGSPGFKAV	YQEIEKGGCG	DPGIPAYGKR	TGSSFLHGDT
	251		LVGERVITCQ		SCVFSCFFNF	TASSGIILSP
10	301	NYPEEYGNNM	NCVWLIISEP	GSRIHLIFND	FDVEPQFDFL	AVKDDGISDI
•	351	TVLGTFSGNE	VPSQLASSGH	IVRLEFQSDH	STTGRGFNIT	YTTFGQNECH
	401	DPGIPINGRR	FGDRFLLGSS	VSFHCDDGFV	KTQGSESITC	ILQDGNVVWS
	451		GGHLTASSGV		KDSLHCEWII	
	501	TFDRFQTEVN	YDTLEVRDGP	ASSSPLIGEY		
15	551	TTDNSRSSIG	FLIHYESVTL	ESDSCLDPGI	PVNGHRHGGD	FGIRSTVTFS
	601	CDPGYTLSDD	EPLVCERNHQ	WNHALPSCDA	LCGGYIQGKS	GTVLSPGFPD
	651	FYPNSLNCTW	TIEVSHGKGV	_	SSHDYLLITE	
÷.	701	LTGSVLPHTI	KAGLFGNFTA			EYDLEPCDDP
	751	GVPAFSRRIG	FHFGVGDSLT		GATKLTCLGG	
20	801		KGNEGTLLSP		ECIYKIETEA	
	851			PLGTFTKNEL		
	901			DPGIPNYGYR		
	951			PLPSCIAECG		
	1001			HFIVFDTEMA		
25	1051			DSDFFISKSG		
	1101			PGYQLQGQAK		
	1151			QPYPPGKECD		
	1201			GSYQGSQAPE		
	1251			NIMNGTRVGT		
30	1301			PSCNAPCGGQ		
	1351			YFQTALNDLA		
	1401			SGASARGFHF		TQCSSVPEPR
	1451			PGYLLQGSTA	-	_
	1501			PEPYGNNLNC		
35	1551			LGSFSGTTVP		
	1601			ALPSNSIKIG		
	1651			LCIATCGGTL		
	1701			FSTEANHDFL		
40	1751			SQNRQGFKLA		
40	1801			LIGHPVLTCQ		
	1851			LKDCIWLITV.		
	1901			NTALETAYSS		
	1951		LVKTENSMWC	LLQCCPTPCF	QLKFLDSAEG	VYDSFALEAS
	2001	VSCGPFFV*				

100	56-3V3 Pr	otein sequen	ice	1784 AA		
	1	MEATKTLSGI	WNNINHVTSE	EDTFIMYLGK	PWLQVKIQVS	QGGVALVSDM
5	51	CPDPGIPENG	RRAGSDFRVG	ANVQFSCEDN	YVLQGSKSIT	CQRVTETLAA
	101	WSDHRPICRA	RTCGSNLRGP	SGVITSPNYP	VQYEDNAHCV	WVITTTDPDK
	151	VIKLAFEEFE.		GDAGKVGDTR	SVLYVLTGSS	
	201	OMWITHLOSDD	SIGSPGFKAV	YQEIEKGGCG	DPGIPAYGKR	
2.65	' 251	LTFECPAAFE	LVGERVITCQ	QNNQWSGNKP	SCVFSCFFNF	TASSGIILSP
10	301	NYPEEYGNNM	NCVWLIISEP	GSRIHLIFND	FDVEPQFDFL	AVKDDGISDI
	351	TVLGTFSGNE	VPSQLASSGH	IVRLEFQSDH	STTGRGFNIT	YTTFGQNECH
	401	DPGIPINGRR	FGDRFLLGSS	VSFHCDDGFV	KTQGSESITC	ILQDGNVVWS
	451	STVPRCEAPC	GGHLTASSGV	ILPPGWPGYY	KDSLHCEWII	EAKPGHSIKI
	501	TFDRFQTEVN	YDTLEVRDGP	ASSSPLIGEY	HGTQAPQFLI	STGNFMYLLF
15	551	TTDNSRSSIG	FLIHYESVTL	ESDSCLDPGI	PVNGHRHGGD	FGIRSTVTFS
	601	CDPGYTLSDD	EPLVCERNHQ	WNHALPSCDA	LCGGYIQGKS	GTVLSPGFPD
	651	FYPNSLNCTW	TIEVSHGKGV	QMIFHTFHLE	SSHDYLLITE	DGSFSEPVAR
	701	LTGSVLPHTI	KAGLFGNFTA	QLRFISDFSI	SYEGFNITES	EYDLEPCDDP
	751	GVPAFSRRIG	FHFGVGDSLT	FSCFLGYRLE	GATKLTCLGG	GRRVWSAPLP
20	801	RCVAECGASV	KGNEGTLLSP	NFPSNYDNNH	ECIYKIETEA	GKGIHLRTRS
	851	FQLFEGDTLK	VYDGKDSSSR	PLGTFTKNEL	LGLILNSTSN	HLWLEFNTNG
	901	SDTDQGFQLT	YTSFDLVKCE	DPGIPNYGYR	IRDEGRETOT	VVLYSCNPGY
	951	AMHGSNTLTC	LSGDRRVWDK	PLPSCIAECG	GQIHAATSGR	ILSPGYPAPY
	1001	DNNLHCTWII	EADPGKTISL	HFIVFDTEMA	HDILKVWDGP	ADSDITTER
25	1051	SGSALPEDIH	STFNSLTLQF	DSDFFISKSG	FSIQESTSIA	ATCNDPGMPQ
	1101		AGDTVTFQCD	PGYQLQGQAK	TTCVQLNNRF.	FWQPDPPTCI
	1151	AACGGNLTGP	AGVILSPNYP	QPYPPGKECD	WKVKVNPDEV	IALIFKSFNM
	1201	EPSYDFLHIY	EGEDSNSPLI	GSYQGSQAPE	RIESSGNSLE	LAFRSDASVG
	1251	LSGFAIEFKE	KPREACEDPG	NIMNGTRVGT	DEKLESTITI	OCDOCIVIDO
30	1301		DGKPSWDQVL	PSCNAPCGGQ	ILCOFGAATO	DITECTORS
	1351	QICLYSITVP	KEFVVFGQFA	YFQTALNDLA	FPLOGIUMON	TOCSSVPEPR
	1401	SGETLPLATS	NOILLRESAK	SGASARGFHF	VIQAVPRISD	
	1451	YGRRIGSEFS	AGSIVREECN	PGYLLQGSTA	THECOSVENAL	CTOTOUTSED
	1501	VVPCSGNFTQ	•	PEPIGNNLNC	TMUTTATEGO	GIQIQVISFA YLHFOSDISV
35	1551	TEQNWDSLEI	HDGGDVTAPR	LGSFSGTTVP		FQCEPGYTLQ
	1601	AAAGFHLEYK	TVGLAACQEP	WILDNOTUTE	OUTHANDARD DE	GFPGSYPNNL
	1651	GRSHISCMPG	TVKKWNIPSP	TCTWICEGIT	PIUGDAHAS	PMIGQFSGTD
	1701	DCTWRISLPI	GYGAHIQELN	EDJ PWNUDED	AUD*	THIOTEGED
	1751	LPAALLSTTH	FITTHEISON	SQNRQGFKLA	1 Au	

	5R-3V2 Pr	otein seque	nce	2353 AA		
	1		XSLRLALPSG		SSEEPSVSRA	CGVSADMTAW
5	. 51	RRFQSLLLLL	GLLVLCARLL	TAAKGQNCGG	LVQGPNGTIE	SPGFPHGYPN
	101	YANCTWILLT	GERNRIQLSF	HTFALEEDFD	ILSVYDGQPQ	QGNLKVRLSG
	151	FQLPSSIVST	· GSILTLWFTT	DFAVSAQGFK	ALYEVLPSHT	CGNPGEILKG
	201	VLHGTRFNIG	DXIRYSCLPG	YILEGHAILT	CIVSPGNGAS	WDFPAPFCRA
• .	251	EGACGGTLRG	TSSSISSPHF	PSEYENNADC	TWTILAEPGD	TIALVFTDFQ
10	301	LEEGYDFLEI	SGTEAPSIWL	TGMNLPSPVI	SSKNWLRLHF	TSDSNHRRKG
	351	FNAQFQVKKA.	IELKSRGVKM	LPSKDGSHKN	SVLSQGGVAL	VSDMCPDPGI
	401		FRVGANVQFS			
	451		LRGPSGVITS			
	501	EEFELERGYD	TLTVGDAGKV	GDTRSVLYVL	TGSSVPDLIV	SMSNQMWLHL
15	551	QSDDSIGSPG	FKAVYQEIEK	GGCGDPGIPA	YGKRTGSSFL	HGDXLTFECP
	601	AAFELVGERV	·ITCQQNNQWS	GNKPSCVFSC	FFNFTASSGI	ILSPNYPEEY
	651	GNNMNCVWLI	ISEPGSRIHL	IFNDFDVEPQ	FDFLAVKDDG	ISDITVLGTF
٠.	701	SGNEVPSQLA	SSGHIVRLEF	QSDHSTTGRG	XNITYTTFGQ	NECHDPGIPI
	751	NGRRFGDRFL.	LGSSVSFHCD	DGFVKTQGSE	SITCILODGN	VVWSSTVPRC
20	801	EAPCGGHLTA	SSGVILPPGW	PGYYKDSLHC	EWIIEAKPGH	SIKITFDRFQ
•	851	TEVNYDTLEV	RDGPASSSPL	IGEYHGTQAP	QFLISTGNFM	YLLFTTDNSR
	901	SSIGFLIHYE	SVTLESDSCL	DPGIPVNGHR	HGGDFGIRST	VTFSCDPGYT
	951		RNHQWNHALP			
	1001	NCTWTIEVSH	GKGVQMIFHT	FHLESSHDYL	LITEDGSFSE	PVARLTGSVL
25	1051	PHTIKAGLFG	NFTAQLRFIS	DFSISYEGFN	ITFSEYDLEP	CDDPGVPAFS
•	1101	RRIGFHFGVG	DSLTFSCFLG	YRLEGATKLT	CLGGGRRVWS	APLPRCVAEC
	1151	GASVKGNEGT	LLSPNFPSNY	DNNHECIYKI	ETEAGKGIHL	RTRSFQLFEG
	1201		SSSRPLGTFT			
•	1251		VKCEDPGIPN			
30	1301		VWDKPLPSCI	_		
	1351		TISLHFIVFD			
	1401		TLQFDSDFFI			
	1451		FQCDPGYQLQ			
25	1501		PNYPQPYPPG			
35	1551		SPLIGSYQGS			
	1601		FDPGNIMNGT			
	1651		DQVLPSCNAP			
	1701		GQFAYFQTAL			
40	1751		FSAKSGASAR			
40	1801		FECNPGYLLQ			
	1851		SPGYPEPYGN			
	1901		TAPRLGSFSG			
	1951		CQEPALPSNS			_
15	2001		YPSPLCIATC			
45	2051		QFLNFSTEAN			
	2101		YSDHSQNRQG			
	. 2151		PGYILIGHPV			
	2201	GTIISPGFPD	EYPILKDCIW	PACCUMOUT	TINETPLOTE	WANDITWAND
50	2251		VFSGNTALET			
50	2301		SMWCLLQCCP	TECTOTKETD	SALGVYDSFA	LEAS VSCGPF
	2351	FV*				

	PROTEIN SEQUENCE 5R23V2					•
	LOCUS 5R23V2.P		A PROT	UE	DATED 05/	11/101
5	DEFINITION -					
,	ACCESSION -					
	KEYWORDS -					
	SOURCE -	•				
	FEATURES From	To/Span	Description	ac		
10		2307	851 to 777	11 of 5R23V2	!translated	i)
		2301	031 00	2 02 01		
•	ORIGIN ? 1 MTAWRRFQSL	TITICITUTO	API.I.TAAKGO	NCCCT.VOCPN	GTIESPGEPH	GYPNYANCTW
	61 IIITGERNRI	DIPPOPULATO	EDEDITIEUVD	COPOCKICKY	RISGFOLPSS	IVSTGSILTL
	121 WFTTDFAVSA	OCEANI AEM	DOUTCONDGE	TI.KCVI.HCTR	FNIGDXIRYS	CLPGYTLEGH
15	181 AILTCIVSPG	VGE VALIEVE	PODRECACCO	TI DOTECTE	COHEDSEALM	NADCTWTTLA
15	181 AILTCIVSPG 241 EPGDTIALVF	NGASWOFPAP	TCRMEGACOG	CTWI TOMNID	COUTESKNWT.	BI.HETSDSNH
	241 EPGDTIALVE	TOFOLERGIO	FLEISGIERF	STADIGMMDE	CANTACOWOD	DECIDENCED
	301 RRKGFNAQFQ	VKKAIELKSK	GVKMLPSKDG	SHKNSVLSQG	GAMPACONCE	CCCNT DCDCC
	361 AGSDFRVGAN	VQFSCEDNYV	LOGSKSITCO	RVILILAAWS	DUKLICKNYI	CGSMTVGESG
~ ^	421 VITSPNYPVQ	YEDNAHCVWV	ITTTDPDKVI	KLAXEEFELE	KGIDITIVGD	AGKVGDIKSV
20	481 LXVLTGSSVP	DLIVSMSNQM	WLHLQSDDSI	GSPGFKAVYQ	EIERGGCGDP	GIPAIGARIG
	541 SSFLHGDXLT	FECPAAFELV	GERVITCQQN	NOWSGNKPSC	VESCEENFTA	SSGIILSPNI
	601 PEEYGNNMNC	VWLIISEPGS	RIHLIFNDFD	VEPQFDFLAV	KDDG1SD1TV	LGTFSGNEVP
	661 SQLASSGHIV	RLEFQSDHST	TGRGXNITYT	TFGQNECHDP	GIPINGRRFG	DRFLLGSSVS
	721 FRCDDGFVKT	OGSESITCIL	ODGNVVWSST	VPRCEAPCGG	HLTASSGVIL	PPGWPGYYKD
25	781 STRCEWITEA	KPGHSIKITE	DRFOTEVNYD	TLEVRDGPAS	SSPLIGEYHG	TQAPQFLIST
-	941 CNEMVILLETT	DNSRSSTGFL	IHYESVTLES	DSCLDPGIPV	NGHRHGGDFG	IRSTVTFSCD
	· 901 PCYTTSDDEP	LVCERNHOWN	HALPSCDALC	GGYIQGKSGT	VLSPGFPDFY	PNSLNCTWTI
	961 EACHCKCAOW	TEHTEHLESS	HDYLLITEDG	SFSEPVARLT	GSVLPHTIKA	GLXGNFTAQL
	1021 BETSDESTSY	EGENTTESEY	DIEPCDDPGV	PAFSRRIGFH	FGVGDSLTFS	CFLGYRLEGA
30	1081 TRITCIGGGR	RVWSAPTIPRC	VAECGASVKG	NEGTLLSPNF	PSNYDNNHEC	IYKIETEAGK
50	1141 CTULDTDSEC	I.FEGDTI.KVV	DCKDSSSRPL	GTFTKNELLG	LILNSTSNHL	WLEFNTNGSD
•	1201 7006501777	SEDIVECEDE	GIPNYGYRIR	DEGHFTDTVV	LYSCNPGYAM	HGSNTLTCLS
	1261 GDRRVWDKPL	PSCTAECGGO	THAATSGRIL	SPGYPAPYDN	NLHCTWIIEA	DPGKTISLHF
	1321 IVFDTEMAHD	TIKVWDGPVD	SDILLKEWSG	SALPEDIHST	FNSLTLQFDS	DFFISKSGFS
35	1381 IQFSTSIAAT	CNUDGMEONG	TRYGDSREAG	DTVTFOCDPG	YOLOGOAKIT	CVQLNNRFFW
55	1441 QPDPPTCIAA	CCCNLTCPAG	VILSPNYPOP	YPPGKECDWR	VKVNPDFVIA	LIFKSFNMEP
	1501 SYDFLHIYEG	EDENEDITES	VOCSOAPERT	ESSGNSLELA	FRSDASVGLS	GFAIEFKEKP
	1561 REACFDPGNI	WACABACADA FD2M2EP162	KICSTITYOC	DSGYKILDPS	SITCVIGADG	KPSWDOVLPS
	1621 CNAPCGGQYT	CCCCUTCDI	VDHNYTAGOT	CLYSITVPKE	FVVFGOFAYE	OTALNDLAEL
40	1681 FDGTHAQARL	GSEGAATSEN	EMILITAGET	TLLBESAKSG	ASARGEHEVY	OAVPRTSDTO
40	1741 CSSVPEPRYG	PERCEESS	EINERGIBLE	VI.I.OGSTATH	COSVENALAO	WNDTTPSCVV
	1801 PCSGNFTQRR	KKIGSEFSAG	DACKET MCIR	TIDDOGGIADII.	OTOVISEATE	ONWINGTHE
	1801 PCSGNFTQRR 1861 GGDVTAPRLG	GTILSPGIPE	PIGNNENCIW	VIIAIFGOGI	ACERT.EVETU	GLARCOEPAL
	1861 GGDVTAPRLG	SESGTIVEAL	PN212MOTIT	ULOSOTSAVV	AGE HEIGHT C	TATCCCTLST
15	1921 PSNSIKIGDR	YMVNDVLSFQ	CERCLIFFORK	CAUTOFINES	WWWIEDERC	ONCOVERSOM
45	1981 LGGVILSPGF	PGSYPNNLDC	TWKISLPIGY	PARTOL THE P	TOWNSDED	DDEONCYMIN
	2041 IGQFSGTDLP	AALLSTTHET	LIHFYSDHSQ	NKQGFKLAYQ	WIFTONCEDE	LEL GROTHIN
	2101 SDYSVGQSVS	FECYPGYILI	GHPVLTCQHG	TUKNMUILEL	ACDAPCGINV	TOUNGITION
	2161 GFPDEYPILK	DCIWLITVPP	GHGVYINFTL	LQTEAVNDYI	AAMDGADONS	CCCDMDCECT
	2221 ALETAYSSTN	QVLLKFHSDF	SNGGFFVLNF	HCOLIFIED A	KTENSMWCLL	OCCLIRCION
50	2281 KFLDSAEGVY	DSFALEASVS	CGPFFV*			

::50	;	5R2 OC147 PROTEIN					
5	LOCUS TO DEFINITION	-	. 347 AJ	A PROT	O:	PDATED 0	5/11/101
10	FEATURES ' Peptide ORIGIN ?	1	To/Span 347		01 of 5r2_0		·
15	61 III 121 WFT 181 AII 241 EPG	TGERNRI TTDFAVSA LTCIVSPG GDTIALVF	QLSFHTFALE QGFKALYEVL NGASWDFPAP TDFQLEEGYD	EDFDILSVYD PSHTCGNPGE FCRAEGACGG FLEISGTEAP	GQPQQGNLKV ILKGVLHGTR TLRGTSSSIS	RLSGFQLPS: FNIGDKIRY: SPHFPSEYEI SPVISSKNWI	GYPNYANCTW SIVSTGSILTL CLPGYILEGH NADCTWTILA RLHFTSDSNH

1 000	*	5R2 AW PROTEIN			
5	DEFINITION - ACCESSION - KEYWORDS -	W_PRO 372 A	A PROT	UPDATED	05/11/101
10	Peptide ORIGIN ?	om To/Span 1 372 QSL LLLLGLLVLC	ARLLTAAKGO N	of 5r2_aw (trans	GFPH GYPNYANCTW
15	121 WFTTDFA 181 AILTCIV 241 EPGDTIA 301 RRKGFNA	VSA. QGFKALYEVL SPG NGASWDFPAP LVF TDFQLEEGYD QFQ VKKAIELKSR	PSHTCGNPGE I FCRAEGACGG T FLEISGTEAP S	QPQQGNLKV RLSGFQ: LKGVLHGTR FNIGDK: LRGTSSSIS SPHFPS: IWLTGMNLP SPVISS: HKNSVWHQQ EFSKCR	IRYS CLPGYILEGH EYEN NADCTWTILA KNWL RLHFTSDSNH
20	361 LTASGNL	QFD N*			

1/3

FIGURE 1.

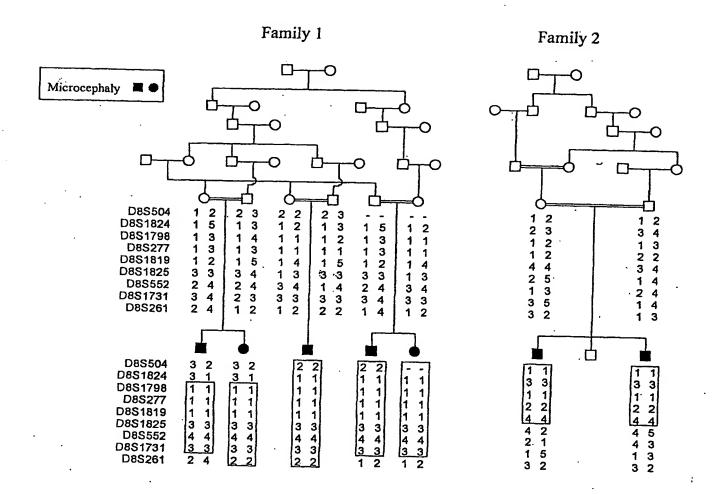
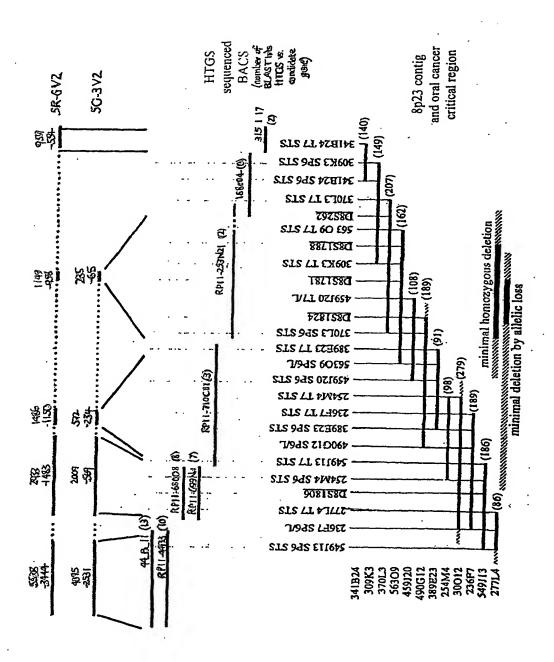
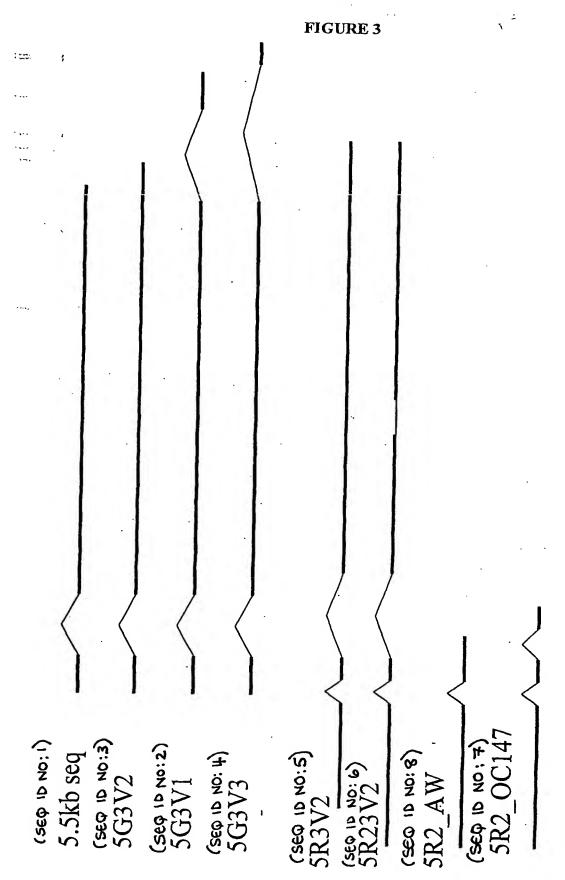


FIGURE 2



SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

n anal Application No PCT/GB 01/02240

A. CLASSIF IPC 7	C12N15/12 C12N15/85 C12N15/ C12Q1/68 G01N33/577 A61K31/		(16/18 (67/027			
According to	International Patent Classification (IPC) or to both national classification	cation and IPC				
B. FIELDS			· · · · · · · · · · · · · · · · · · ·			
Minimum do IPC 7	cumentation searched (classification system tollowed by classifica C12N C07K C12Q G01N A61K A01	ion symbols) K				
Documentati	ion searched other than minimum documentation to the extent that	such documents are included in the fields	searched			
Electronic da	ata base consulted during the international search (name of data b	ase and, where practical, search terms use	d)			
WPI Da	WPI Data, PAJ, CAB Data, STRAND, BIOSIS, EPO-Internal					
· ·						
	ENTS CONSIDERED TO BE RELEVANT	November 200	Delevent to all-lands			
Category °	Citation of document, with indication, where appropriate, of the n	elevant passages	Relevant to claim No.			
X	DATABASE EMBL SEQUENCE DATABASE Hinxton, UK; 21 October 1999 (19 NCI-CGAP: "xd71c12.x1 Soares_NFL Homo sapiens cDNA clone IMAGE:26 mRNA sequence; EST" XP002175139 EMBL:AW104197, Comparison of Acc AW104197 and SEQ ID No. 8; abstract	99-10-21) _T_GBC_S1 03062 3'	1-6			
X Funi	her documents are listed in the continuation of box C.	Patent family members are liste	d in annex.			
*To later document published after the International filing date of priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention cannot be considered novel or cannot be considered to be after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention or cited to understand the principle or theory underlying the invention or particular relevance; the ctaimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone which is cited to establish the publication date of another citation or other special reason (as specified). "O' document referring to an oral disclosure, use, exhibition or other means." "P' document published prior to the international filing date but later than the priority date claimed." "It ater document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention cannot be considered novel or cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art."						
	actual completion of the international search	Date of mailing of the International s	earch report			
1	7 August 2001	29/08/2001				
Name and r	mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Hornig, H				

INTERNATIONAL SEARCH REPORT

In anal Application No
PCT/GB 01/02240

C.(Continu:	ation) DOCUMENTS CONSIDERED TO BE RELEVANT	rc1/88 01/02240
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL SEQUENCE DATABSE 'Online! Hinxton, UK; 3 March 2000 (2000-03-03) L. HILLIER ET AL.: "zj96c05.s1 Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA clone; EST" XP002175140 EMBL:AA705177, Comparison of Accession no. AA705177 and SEQ ID No.2; abstract	1-6
x	DATABASE EMBL SEQUENCE DATABASE 'Online! Hinxton, UK; 2 January 2000 (2000-01-02) K. KYUNG ET AL.: "Homo sapiens BAC clone RP11-221H10 from 8, complete sequence; HTG" XP002175141 EMBL:AC019176, Comparison of Accession no. AC019176 from position 13832-14773 and SEQ ID No. 6; abstract	1-6
A .	SUN PAUL C ET AL: "Homozygous deletions define a region of 8p23.2 containing a putative tumor suppressor gene." GENOMICS, vol. 62, no. 2, 1 December 1999 (1999-12-01), pages 184-188, XP002175136 ISSN: 0888-7543 cited in the application the whole document	·
A	ISHWAD CHANDRAMOHAN S ET AL: "Frequent allelic loss and homozygous deletion in chromosome band 8p23 in oral cancer." INTERNATIONAL JOURNAL OF CANCER, vol. 80, no. 1, 5 January 1999 (1999-01-05), pages 25-31, XP002175137 ISSN: 0020-7136 the whole document	
4	SUNWOO JOHN B ET AL: "Localization of a putative tumor suppressor gene in the sub-telomeric region of chromosome 8p." ONCOGENE, vol. 18, no. 16, 22 April 1999 (1999-04-22), pages 2651-2655, XP001015856 ISSN: 0950-9232 the whole document	·

INTERNATIONAL SEARCH REPORT

n stional Application No PCT/GB 01/02240

C.(Continue	tion) DOCUMENTS CONSIDERED TO BE RELEVANT	
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	SUN PAUL C ET AL: "Transcript map of the 8p23 putative tumor suppressor region." GENOMICS, vol. 75, no. 1-3, July 2001 (2001-07), pages 17-25, XP002175138 ISSN: 0888-7543 the whole document	

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claim 20 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claim(s) 18 and 19 (as far as in vivo methods are concerned) are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.